LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 7
marzo 1969
anno II
volume II

L'evoluzione del cervello e la talidomide

In questo articolo l'autore sostiene che la sua teoria della «evoluzione stratificata» del cervello umano avrebbe consentito di prevedere gli effetti focomelici della talidomide.

di Renato Balbi

L'azione teratogena che la talidomide esercita nell'uomo è nota a tutti: se questo farmaco viene somministrato alle gestanti in un determinato periodo della gravidanza, i figli presentano, alla nascita, gravi malformazioni. Viceversa, del tutto sconosciuta è la causa di questa azione, nonostante gli esperimenti che sono stati effettuati negli ultimi tempi da eminenti studiosi di discipline diverse (fisiologi, farmacologi, clinici e genetisti), i quali ne hanno riferito al Tribunale tedesco chiamato a giudicare i « padri » della talidomide.

Tali esperimenti hanno consentito, comunque, di approfondire le conoscenze relative agli effetti determinati dalla talidomide, oltre che nell'uomo, in animali appartenenti a specie diverse. Nel soggetto umano adulto, il farmaco si limita, di solito, a esercitare una blanda azione sedativa sul sistema nervoso (favorisce il sonno): è significativo che la somministrazione della talidomide modifichi il tracciato elettroencefalografico in modo da renderlo perfettamente simile a quello del sonno fisiologico. Si sono peraltro registrati dei casi – percentualmente poco numerosi – nei quali, a distanza di tempo, il farmaco ha provocato una polinevrite di tipo tossico.

Venendo agli animali, gli esperimenti del professor L.H. Schmidt hanno dimostrato esaurientemente (dopo i primi significativi risultati ottenuti negli Stati Uniti da C.S. Delahunt e collaboratori, e da A.G. Hendrickx e collaboratori) che la talidomide esercita nelle scimmie un'azione teratogena simile a quella esercitata nell'uomo: in altri termini, se una scimmia femmina ingerisce il farmaco durante la gravidanza, i suoi piccoli nasceranno con malformazioni analoghe a quelle riscontrate, in casi simili, nei bambini. Ma questo fenomeno riguarda esclusivamente le scimmie: negli altri animali l'azione focomelica della talidomide non ha luogo.

Fin qui i principali risultati degli esperimenti che sono stati illustrati al processo di Alsdorf. Questo processo, come è noto, ha sollevato interrogativi che oltrepassano i limiti del dibattito scientifico e investono un problema di ordine morale: coloro che hanno sintetizzato la talidomide e l'hanno introdotta in clinica umana debbono essere considerati responsabili di quanto è accaduto? oppure si tratta dell'inevitabile scotto che la scienza deve pagare per il proprio progresso?

Dico subito che la risposta da dare all'uno come all'altro interrogativo non può essere che negativa. E cioè: 1) i « padri » della talidomide non possono essere accusati di negligenza criminale per non avere previsto gli effetti teratogeni che il farmaco avrebbe determinato nell'uomo; 2) la previsione di tali effetti era possibile, e era quindi possibile scongiurare l'introduzione del farmaco in clinica umana, con le sue dolorose conseguenze.

Le due affermazioni sembrano contraddittorie; ma non lo sono. E' vero che i ricercatori si astennero dall'effettuare i soli esperimenti – quelli sulle scimmie – che avrebbero potuto metterli in guardia circa l'azione teratogena della talidomide; ma è anche vero che ciò non dipese da una loro criminale leggerezza. Mi propongo di chiarirne il perché.

Ogniqualvolta si studiano gli effetti provocati da un farmaco, è necessario distinguere tra la *sede* sulla quale la sostanza agisce con il suo effetto farmacologico (inibizione, eccitazione) e il *meccanismo* biochimico che costituisce la base di tale effetto farmacologico.

Mi spiegherò con un esempio. Molto prima che fosse nota l'azione anticolinergica del curaro, Claude Bernard aveva dimostrato che questa sostanza agisce sulla placca motrice: la blocca, e in tal modo provoca nel soggetto una paralisi muscolare. Quanto al meccanismo biochimico del fenomeno, esso è stato chiarito in tempi piú vicini a noi. Si è accertato che il curaro ha un'azione competitrice con l'acetilcolina, ossia con il mediatore chimico che trasmette l'impulso nervoso alle placche motrici. Data la sua affinità con i recettori delle placche motrici, il curaro li occupa: di conseguenza, l'acetilcolina liberata dal-

l'impulso nervoso, trovando occupati i recettori, non è in grado di trasmettere l'impulso stesso (press'a poco come una chiave non è in grado di aprire una porta se la serratura è occupata da un'altra chiave). Tuttavia, come ho accennato piú sopra, anche prima che si conoscesse l'azione anticolinergica del curaro, era nota la sede sulla quale esso agisce (la placca motrice), e era noto il suo effetto farmacologico (la paralisi muscolare).

Anche l'atropina è una sostanza anticolinergica (competitrice con l'acetilcolina). L'atropina, però, a differenza del curaro, possiede affinità con i recettori parasimpatici, i quali sono anch'essi stimolati dall'acetilcolina. L'atropina perciò non provoca paralisi muscolare, ma dà luogo a un deficit del parasimpatico. Ecco dunque due sostanze che hanno una stessa interferenza biochimica (azione competitrice con l'acetilcolina) ma provocano un effetto farmacologico diverso. È dunque evidente che, oltre al tipo di interferenza biochimica posseduta da un farmaco, occorre, per chiarirne l'effetto, conoscere le strutture interessate dall'azione del farmaco stesso, e accertare se queste strutture ne sono stimolate o inibite.

Tornando alla talidomide, sono stati effettuati studi assai pregevoli per stabilire l'interferenza del farmaco in processi enzimatici: ricorderò, tra gli altri, gli esperimenti di R.M. Boergstrom, dai quali si deduce che la talidomide agisce essenzialmente su sistemi enzimatici che interferiscono con il metabolismo degli aminoacidi; lo stesso Boergstrom ha inoltre dimostrato che la talidomide modifica l'elettroencefalogramma anche nei roditori. Ma tutto ciò non ci aiuta a chiarire *perché* la talidomide esercita un'azione teratogena. In questo articolo, io mi propongo di tralasciare la questione della base biochimica, che pure costituisce il substrato intimo dell'azione del farmaco; mi riprometto piuttosto – cosí come fece Claude Bernard per il curaro – di individuare quale sia la sede d'azione della talidomide, e dedurne quindi gli effetti.

Non avendo avuto modo di procedere a esperimenti personali, cercherò di interpretare i dati esistenti, ossia i dati, cui accennavo in principio, emersi al processo di Alsdorf.

Mi sia concesso rilevare che da tempo avevo raggiunto la convinzione che la talidomide esercita, nelle scimmie, la stessa azione teratogena riscontrata nell'uomo. E avevo espresso tale convinzione in un lavoro pubblicato prima che fossero resi noti i risultati degli esperimenti condotti con la talidomide sulle scimmie. Dico questo perché è appunto la teoria illustrata nel mio lavoro *L'evoluzione stratificata*, che, a mio avviso, consente – tra l'altro – di chiarire la causa dell'azione teratogena esercitata dal farmaco.

Un altro dato emerso al processo di Alsdorf merita di essere sottolineato: il professor W. Lenz, osservando i bambini focomelici nati da donne che avevano fatto uso della talidomide durante la gravidanza, era in grado di stabilire esattamente il periodo in cui la madre era ricorsa al farmaco (il che, come si vedrà piú oltre, costituisce un'ulteriore, importante conferma della validità della mia ipotesi). Aggiungerò che gli esperimenti di Schmidt confermano la teoria di Lenz, secondo la quale la talidomide provoca i suoi mostruosi effetti solo se presa dalla gestante in una particolare fase della gravidanza.

Mi sembra dunque che sia lecito trarre, da quanto ho ricordato finora, due conclusioni: 1) la talidomide agisce, tanto nell'uomo adulto quanto nell'embrione umano, su strutture che non sono presenti nei roditori e nei carnivori; 2) la talidomide possiede effetto teratogeno se la gestante fa uso del farmaco nell'epoca in cui nell'embrione avviene la organizzazione di queste strutture.

È presumibile che l'affinità della talidomide con tali strutture non sia una affinità assoluta, poiché, in seguito alla somministrazione del farmaco, sono state dimostrate interferenze cliniche (diminuzione dei movimenti spontanei), biochimiche ed elettroencefalografiche anche nei roditori. In questi animali, nondimeno, tali interferenze non si traducono né in effetti terapeutici né in azione teratogena o comunque tossica. A questo proposito, è interessante ricordare che la talidomide non è stata introdotta negli Stati Uniti d'America perché, mancando nei roditori l'effetto terapeutico riscontrato nell'uomo, gli organi responsabili ritennero priva di valore la documentazione che tendeva a dimostrare l'innocuità del farmaco (documentazione che era basata, per lo appunto, su esperimenti condotti sui roditori). Fu questa, possiamo dirlo, una intuizione geniale. Ma era soltanto una intuizione; e non consentiva di comprendere il meccanismo d'azione della talidomide.

Poiché, come si è detto, il farmaco non possiede un'affinità specifica con i tessuti umani, l'unica ipotesi attendibile mi sembra essere la seguente: che la talidomide possieda la massima affinità con strutture presenti nell'uomo (e nei primati), e che l'azione esercitata dal farmaco su tali strutture sia la causa degli effetti – terapeutici, tossici e teratogeni – che si sono riscontrati.

Ora, quali sono queste strutture? Per poter rispondere, dovrò prima esporre brevemente i principi sui quali si fonda la teoria alla quale ho accennato, la teoria che ho definito dell'« evoluzione stratificata ». Sono convinto che questa teoria possa non solamente far comprendere il meccanismo d'azione degli psicofarmaci, e dei farmaci teratogeni – come, appunto, la talidomide –, ma anche chiarire le stesse basi fisiologiche della nostra vita psichica, illuminando fenomeni considerati fin qui oscuri dalla scienza (dalla ipnosi alle percezioni incoscienti, dalle doppie personalità alla genesi e al significato dei

Azione anticolinergica del curaro e dell'atropina nella sinapsi (zona di contatto) tra una fibra motrice o una fibra parasimpatica e il rispettivo recettore (effettore specifico). Il curaro blocca l'acetilcolina che, liberata dalla fibra motrice dovrebbe stimolare l'effettore muscolare. L'atropina blocca l'acetilcolina che, liberata dalle fibre parasimpatiche, dovrebbe stimolare l'effettore parasimpatico. L'azione di blocco avviene perché sia il curaro sia l'atropina agiscono da sostanze competitrici nei riguardi dell'acetilcolina.

SCUOLA NEUROLOGICA INGLESE HUGHLINGS JACKSON	SCUOLA NEUROFISIOLOGICA RUSSA IVAN P. PAVLOV	NEUROANATOMIA COMPARATA EDINGER, KAPPERS, HERRICK	PSICHIATRIA PSICOANALITICA SIGMUND FREUD	SINTESI
LIVELLO SUPERIORE	SISTEMA DEL SECONDO SEGNALE		SUPER EGO	ASTRAZIONE DISCRIMINAZIONE SIMBOLIZZAZIONE COMUNICAZIONE
LIVELLO MEDIO	RIFLESSI CONDIZIONATI		EGO	COMPORTAMENTO ADATTIVO ACQUISITO
LIVELLO INFERIORE	RIFLESSI NON CONDIZIONATI		ID	COMPORTAMENTO INNATO STEREOTIPATO

Tavola comparativa dell'evoluzione cerebrale secondo ipotesi di varie scuole neuropsichiatriche. Le aree tratteggiate nel primo cervello in alto sono le piú recenti, filogeneticamente e ontogeneticamente. Meno recenti sono le aree tratteggiate nel cervello di mezzo, mentre nell'ultimo sono le piú antiche. I livelli di Jackson corrispondono ai superstrati dell'evoluzione stratificata.

sogni) e consentendo alla neurochirurgia di ottenere in futuro, mediante interventi stereotassici, il recupero di funzioni perdute. Ma in questa sede mi limiterò a considerare il problema del meccanismo d'azione degli psicofarmaci in genere, e della talidomide in particolare.

Cosí come esso è oggi, l'encefalo umano è il risultato dell'evoluzione. Per evoluzione intendo sia l'evoluzione filogenetica (della specie umana), dalla comparsa della vita sulla terra fino all'uomo attuale, sia l'evoluzione ontogenetica (dell'individuo), dal momento del concepimento fino alla pubertà, vale a dire fino all'epoca in cui l'individuo diviene in grado di procreare un altro individuo della stessa specie.

L'importanza dell'evoluzione per gli studi di neuropsichiatria fu già riconosciuta da H. Jackson; e su questa importanza concordano oggi tutti i piú eminenti studiosi di neurologia, di psichiatria e di psicologia. Poiché Jackson era un seguace di Spencer (secondo il quale l'evoluzione consisterebbe nel passaggio dal piú semplice al piú complesso), egli ritenne che l'evoluzione del sistema nervoso consistesse nel continuo sopraggiungere di nuove e piú complesse funzioni, che lo stesso Jackson inquadrò in una rigida gerarchia. Una volta subentrata la nuova funzione, questa non si sarebbe limitata a esistere, ma avrebbe preso sotto il proprio controllo – sempre secondo Jackson – le funzioni comparse in una fase piú antica, tanto dell'evoluzione ontogenetica quanto dell'evoluzione filogenetica.

Ipotizzando l'esistenza di sistemi neuronici funzionanti in piani diversi, Jackson individuò tre principali livelli: 1) un primo livello, nel quale si realizzano i fenomeni elementari riflessi; 2) un secondo livello, dove hanno luogo gli atti automatici; 3) un terzo livello, nel quale si verificano gli atti volontari.

Pur sostenendo che questi livelli non sono assimilabili a strutture anatomo-fisiologiche, Jackson ritenne che il primo livello potesse essere localizzato nel midollo; il secondo, in corrispondenza delle aree sensoriali e motorie della corteccia; il terzo, infine, nei lobi frontali. Come ho detto, secondo Jackson la evoluzione avrebbe trasformato le funzioni, determinando il passaggio dal piú semplice al piú complesso, dall'automatico al volontario.

Ora, se l'intuizione di Jackson deve ritenersi geniale, la teoria da lui elaborata pecca di semplicismo e quindi non consente di chiarire le leggi che regolano l'attività del sistema nervoso, soprattutto per quanto riguarda la sfera psichica. A differenza delle altre teorie neojacksoniane, la mia ipotesi si rifà, anziché alla concezione di Spencer, a quella di Darwin; o, per dir meglio, a quella parte della concezione darwiniana conosciuta sotto il nome di « trasformismo ». Per Darwin, l'evoluzione consiste nella trasformazione degli esseri piú semplici, vissuti nel periodo arcaico, in esseri via via piú complessi, fino all'uomo civilizzato del nostro tempo: se, in un determinato periodo geologico, una specie esaurisce il proprio potenziale filogenetico (ossia la sua spinta evolutiva), essa rimane sta-

Rappresentazione grafica dell'ipotesi dell'evoluzione stratificata del cervello. Nel corso dell'evoluzione ontogenetica e filogenetica si stratificano nell'encefalo le strutture nervose che determinano il comportamento dell'individuo. Le strutture attive in ciascuna fase evolutiva formano vari livelli che sono raggruppabili in superstrati, il primo dei quali è costituito da 6 livelli, gli altri due da sette. Le strutture dal 1° superstrato hanno per obiettivo la sopravvivenza e sono correlate con la vita vegetativa. Nell'ontogenesi i livelli dal 1° superstrato sono attivi nelle successivi fasi della vita intrauterina, nella filogenesi sono invece attivi nei progenitori dell'uomo vissuti nell'era arcaica, nell'era primaria e all'inizio dell'era secondaria. I livelli del 2° superstrato corrispondono, nell'ontogenesi, al primo anno di vita, nella filogenesi, agli stadi dei mammiferi non umani. Il 2° superstrato è detto anche del piacere, in quanto i vari livelli che si sviluppano sono appunto quelli della vita affettiva. Il 3° superstrato raggruppa i livelli che hanno per obiettivo il bene. Nell'ontogenesi sono attivi nel periodo che va dal 14° mese di vita alla seconda fanciullezza (tra il 10° e il 14° anno di vita); nella filogenesi corrispondono invece al periodo che va dal Pliocene (durante il quale l'antenato dell'uomo apparteneva già al genere *Homo*, ma non ancora alla specie *Homo sapiens*) fino all'era dell'uomo attuale. L'obiettivo « piacere » è sostituito dall'obiettivo « bene » (che va inteso, naturalmente, come quello che la coscienza dell'individuo stabilisce essere « bene »). Durante queste fasi evolutive, i comportamenti sono in rapporto, oltre che con i desideri, anche con gli elementi conoscitivi. Da ultimo viene quello che l'autore chiama livello « extra », al quale appartiene l'individuo che ha superato le fasi evolutive fissate dall'eredità biologica.

zionaria nel corso dei periodi geologici successivi.

Prima di procedere oltre, sarà opportuno precisare che il luogo comune, secondo il quale Darwin avrebbe sostenuto la discendenza dell'uomo dalla scimmia (cosí come la scimmia si presenta oggi) è del tutto arbitrario. Darwin ha semplicemente affermato che l'antenato dell'uomo, se fosse in vita ancor oggi, verrebbe classificato come un quadrumane: ma le scimmie attuali sono molto diverse dalla scimmia-antenato dell'uomo. D'altra parte, nel periodo silurico dell'èra primaria gli unici vertebrati esistenti sulla Terra erano i pesci primitivi. Perciò non si può sfuggire alla conclusione che l'antenato dell'uomo, nel periodo silurico, doveva essere simile a un pesce primitivo. La teoria evoluzionistica non può prescindere dal trasformismo.

Ora, come ho detto, l'ipotesi da me formulata si fonda sull'applicazione in neuropsichiatria del trasformismo. Ed ecco in che consiste questa ipotesi. In ogni successivo stadio dell'evoluzione ha luogo la comparsa di nuove funzioni neuropsichiche, mentre le funzioni già esistenti si modificano; al tempo stesso si formano nuove strutture nervose, correlate con le nuove funzioni o con le funzioni modificate. Le strutture nervose piú antiche, correlate con le funzioni primitive, si stratificano nell'encefalo; esse diventano peraltro inattive, in quanto inibite dagli strati superiori, costituiti, appunto, dalle strutture nuove. E è alle nuove che quelle arcaiche « consegnano » le loro funzioni.

Non sarà inutile ricordare, a questo punto, che le prove della stratificazione sono numerosissime, in neurologia come in psichiatria. In molti casi si è verificato il ritorno all'attività di strutture arcaiche, quando i centri superiori sono stati lesi (e quando, per conseguenza, è cessata l'azione inibitoria che essi esercitavano sulle strutture piú antiche). Per esempio, in condizioni patologiche particolari si sono riscontrate reazioni (come il segno di Babinski, o il riflesso di Magnus) che nell'uomo adulto sano non si verificano, ma che invece erano reazioni normali tanto durante i primi mesi di vita dell'uomo, cioè in un precedente stadio dell'evoluzione ontogenetica, quanto in animali posti, rispetto all'uomo, in gradini piú bassi della scala evolutiva (ma piú o meno simili a un progenitore dell'uomo), cioè in animali situati in un precedente stadio dell'evoluzione filogenetica. La « liberazione » dei centri arcaici, provocata da lesioni dei centri superiori, è stata messa in luce da Jackson, da C.S. Sherrington e da altri studiosi: tuttavia non se ne sono ricavate, a mio avviso, tutte le deduzioni che se ne potevano trarre, e ci si è limitati a farne la base per i tentativi diretti a localizzare una data lesione. Io ritengo, invece, che il fenomeno della « liberazione » richieda un'interpretazione molto piú approfondita. Secondo la mia teoria, la liberazione determina la « riconsegna » delle funzioni ai centri arcaici, e consente in tal modo il compenso funzionale; se questo compenso non si verifica spontaneamente – e questo è un punto capitale – nulla esclude che lo si possa provocare rimuovendo le strutture inibitorie sui centri arcaici: ciò consentirà il recupero di funzioni che oggi si considerano irrimediabilmente perdute.

(È necessario precisare che non tutte le strutture arcaiche persistono indefinitamente nell'encefalo: le strutture correlate con le funzioni piú antiche vengono cancellate dall'evoluzione, onde in nessun caso, per esempio, l'uomo potrebbe tornare a vivere come il pesce suo progenitore. Tuttavia, nel caso delle strutture correlate con le funzioni psichiche, funzioni che sono comparse in epoca relativamente recente, non vi è stata cancellazione, e quindi la riconsegna è sempre possibile.)

È interessante ricordare come il principio della « consegna » delle funzioni psichiche sia già stato sostenuto in psicologia (dagli « psicologi degli strati »: E. Rothacker, H. F. Hoffmann, Ph. Lersh). Ora, secondo la mia teoria, questo principio vale non soltanto per le funzioni psichiche, ma per tutte le funzioni correlate con le strutture presenti nel nevrasse.

Ho accennato prima alla comparsa, durante il processo evolutivo, di nuove funzioni e di nuove strutture nervose. Debbo ora aggiungere che, secondo quanto afferma la legge di Haeckel (che ritengo valida, sia pure con talune limitazioni, anche in neuropsichiatria), l'ontogenesi ricapitola la filogenesi. Sono quindi convinto che le strutture nervose maturano, nel corso dell'ontogenesi, nello stesso ordine in cui sono comparse nella filogenesi: pertanto, in ogni stadio dell'ontogenesi sono attive quelle stesse strutture che erano attive nel corrispondente stadio della filogenesi. L'insieme di queste strutture costituisce quello che io ho chiamato un livello ontofilogenetico. D'altra parte, sebbene in uno stadio omologo dell'ontogenesi e della filogenesi siano attive, come ho detto, le stesse strutture nervose, la maggior parte delle funzioni neuropsichiche sono invece diverse: e ciò perché ogni funzione è la risultante di fattori non solo interni, ma anche esterni. In altre parole, ogni fun-

zione dipende anche dagli stimoli ambientali che agiscono sulle strutture nervose attive e che trasformano cosí le predisposizioni funzionali (con le quali, appunto, sono generalmente correlate le strutture nervose) in disposizioni e dispositivi. Farò un esempio. Il centro di Broca è correlato con la predisposizione per il linguaggio parlato; ma sono le parole percepite a rendere possibile la formulazione del linguaggio (che è una disposizione). Se le parole appartengono alla lingua italiana, il soggetto apprenderà e sarà in grado di parlare l'italiano (dispositivo); se le parole appartengono alla lingua inglese, il soggetto potrà parlare l'inglese (altro dispositivo). È nel dispositivo che si traduce la funzione.

Se si accetta la teoria del neurotropismo funzionante (di Lugaro) – secondo la quale si crea un'associazione tra i neuroni che funzionano in uno stesso periodo –, si perviene alla conclusione che tra le strutture nervose comparse in una stessa epoca (nella filogenesi), come pure tra le strutture nervose maturate in uno stesso periodo di tempo (nell'ontogenesi) si creano associazioni privilegiate. È appunto l'insieme unitario di queste strutture, che io ho chiamato « livello ». Il livello attivo determina il comportamento del soggetto, provocando la sua reazione agli stimoli interni ed esterni. I livelli dei precedenti stadi evolutivi – livelli che sono presenti nell'encefalo, ma che in condizioni normali sono inattivi – costituiscono i « livelli stratificati ». Le strutture correlate con le nuove funzioni comparse in un determinato stadio evolutivo, o con le funzioni modificatesi in quello stadio (strutture che compaiono, come ho detto, nello stesso stadio evolutivo che vede la comparsa delle funzioni nuove o modificate), costituiscono i centri propri del corrispondente livello. Se talune funzioni, nonostante il passaggio da uno stadio all'altro del processo evolutivo, rimangono inalterate, le strutture correlate con tali funzioni, anziché consegnarle a nuovi centri, si integrano nel nuovo livello: sono questi i centri « integrati ». I centri integrati fanno dunque parte di due o piú livelli, onde sono attivi sia quando funziona il livello superiore, sia quando funziona un livello arcaico.

Ho detto piú sopra che la teoria dell'evoluzione stratificata si rifà a Darwin anziché a Spencer. Ciò spiega perché le mie conclusioni sono diverse da quelle di Jackson e degli altri neojacksoniani. In primo luogo, se si ammette il darwinismo, si deve ammettere l'esistenza, nell'encefalo, di tanti livelli quante sono le fasi dell'evoluzione ontofilogenetica. E soltanto in questo modo si possono chiarire soddisfacentemente le leggi che regolano l'attività del sistema nervoso. Per esempio – e ritorniamo cosí al tema di questo scritto, che è la talidomide –, solo ammettendo l'esistenza di centri integrati possiamo comprendere il meccanismo di azione di molti farmaci che agiscono sul sistema nervoso. Secondo la mia ipotesi, tali farmaci hanno una affinità elettiva con i neuroni appartenenti a un determinato livello (o a piú livelli), mentre non la possiedono con i neuroni appartenenti ad altri livelli, sia pure anatomicamente vicini ai primi. Ciò significa che le strutture di un dato livello possiedono alcune caratteristiche biologiche comuni, che invece non sono possedute dai neuroni di livelli diversi. I farmaci dotati di elettività verso queste caratteristiche biologiche colpiranno esclusivamente i neuroni che ne sono in possesso, e cioè eserciteranno un'azione stimolante, o inibitoria, unicamente sulle strutture di un determinato livello. Ora, i centri integrati, anche se attivi nei livelli superiori, possiedono tuttavia le caratteristiche biologiche proprie dei neuroni del loro livello originario: se si accetta questa tesi, la spiegazione del diverso effetto che taluni farmaci provocano nell'uomo normale, nello psicotico e negli animali, diventa semplice.

Prima di esaminare il meccanismo di azione dei farmaci che agiscono sul sistema nervoso centrale, e in particolare il meccanismo d'azione della talidomide, sarà opportuno accennare alle successive fasi dell'evoluzione filogenetica. Si ricorderà qual è la linea dello uomo, cosí come è stata ricostruita dai neodarwiniani. Gli invertebrati dell'èra arcaica si trasformarono in urocordati; seguirono, via via, gli anfiossi, i pesci, gli anfibi, i rettili, i monotremi, i marsupiali, gli insettivori, le scimmie e finalmente l'uomo.

Ora, lo studio ontogenetico del cervello (e in particolare la constatazione dello sviluppo del rinencefalo e del corpo calloso) mi ha portato a concludere che tra gli insettivori e i primati debbono esservi state due tappe intermedie, due anelli: i roditori e i carnivori. Perciò, secondo la teoria dell'evoluzione stratificata, la catena della filogenesi è identica a quella dei neodarwiniani fino agli insettivori: seguono poi i roditori, i carnivori, i primati (non umani) e infine l'uomo.

Ciò non significa, naturalmente, che l'antenato dell'uomo fosse uguale a un rettile, o a un roditore, simile ai rettili o ai roditori di oggi. La maggior parte delle specie animali attuali si sono evolute anch'esse rispetto ai loro antenati vissuti nelle epoche passate; poche sono le specie che si sono fissate in periodi remoti. Bisogna ricordare che la evoluzione presenta molti rami collaterali: in taluni casi il processo evolutivo ha determinato il passaggio da una classe a un'altra, in altri casi, da un ordine all'altro, e in altri ancora si è limitato a provocare una modificazione dei generi e delle specie. Ricorderò un esempio, tra i molti esistenti in letteratura: T. Edinger ha dimostrato che il rinoceronte e l'elefante discendono da uno stesso erbivoro primitivo, vissuto nel terziario superiore: solo che, mentre il rinoceronte si è presto fissato in una forma stabile, l'elefante è andato incontro a ulteriori fasi evolutive.

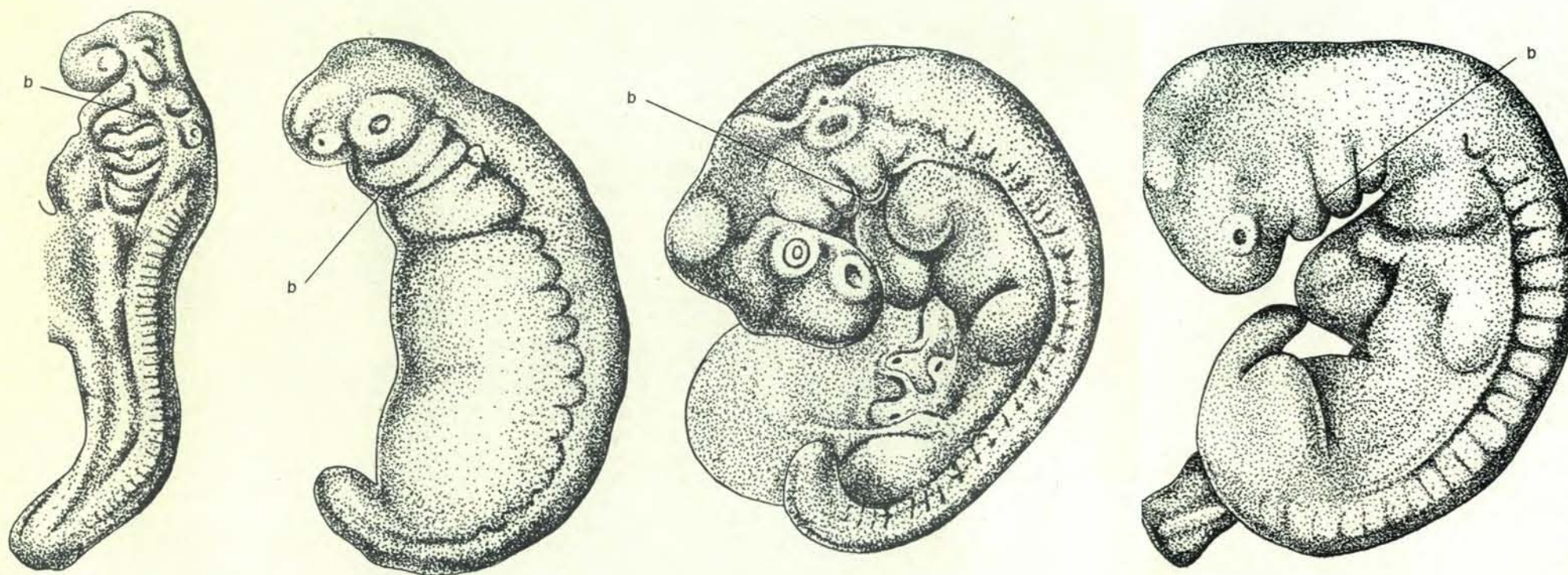
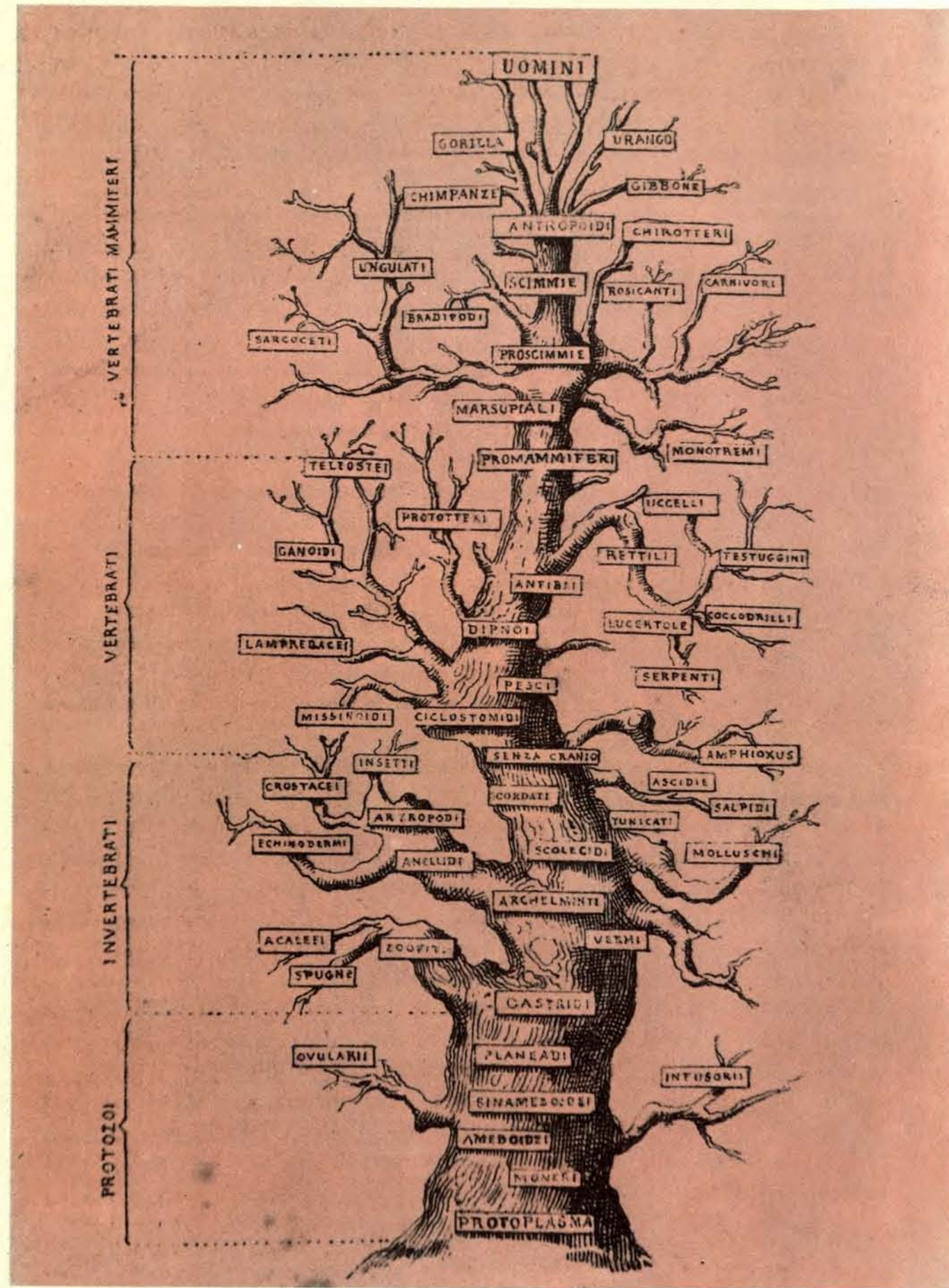
Quando i caratteri essenziali si mantengono, assistiamo alla modificazione del genere o della specie. Quando i caratteri essenziali non si mantengono, i discendenti apparterranno a classi e a ordini diversi. Ciò dipende dal maggiore o minore potenziale filogenetico dei progenitori. Come ricorda H.W. Smith, i vari pesci polmonati del periodo devonico hanno avuto un avvenire evolutivo molto diverso: alcuni di quei pesci subirono soltanto lievi modificazioni, e diedero origine agli attuali dipnoi; altri, viceversa, subirono modificazioni di importanza maggiore, anche se disuguale: una parte di essi diede origine ai pesci teleostei, un'altra parte ai pesci elasmobranchi, una terza parte, infine, ai vertebrati terrestri.

E veniamo ai differenti effetti determinati da taluni farmaci nelle varie specie animali e nell'uomo. Quando il soggetto ha superato lo stadio evolutivo nel quale sono attive le strutture che possiedono affinità con il farmaco, questo limiterà la propria azione ai centri integrati (ossia ai centri che, pur essendo attivi nel livello superiore, conservano le caratteristiche biologiche proprie del livello arcaico). Cosí, i neurolettici, i quali hanno affinità con strutture che nell'uomo normale sono inattive, praticamente in un soggetto sano non provocano alcun effetto, con una sola eccezione: poiché nel livello attivo nell'uomo normale esiste un centro integrato (appartenente in origine a un livello che possedeva affinità con i neurolettici), e cioè il centro correlato con l'affettività legata al mondo esterno, la somministrazione di neurolettici all'uomo normale fa diminuire la tensione emotiva e la reattività. Una somministrazione prolungata giunge a provocare la neurolessia, che è l'indifferenza affettiva agli avvenimenti esterni: si ha una diminuzione dell'iniziativa motoria e un aumento dei fattori intratensivi (dovuti alla persistenza dell'affettività strutturale dell'Io). Se quest'ultima è stimolata, il farmaco potrà esercitare una azione ansiogena.

Viceversa, negli altri mammiferi le strutture interessate dai neurolettici sono attive: su queste strutture il farmaco esercita un'azione inibitoria, cosí che, negli animali citati, i neurolettici provocano numerosi effetti farmacologici: tra i quali la perdita dei riflessi condizionati e l'inibizione della motilità, che può giungere fino alla sindrome catalettica.

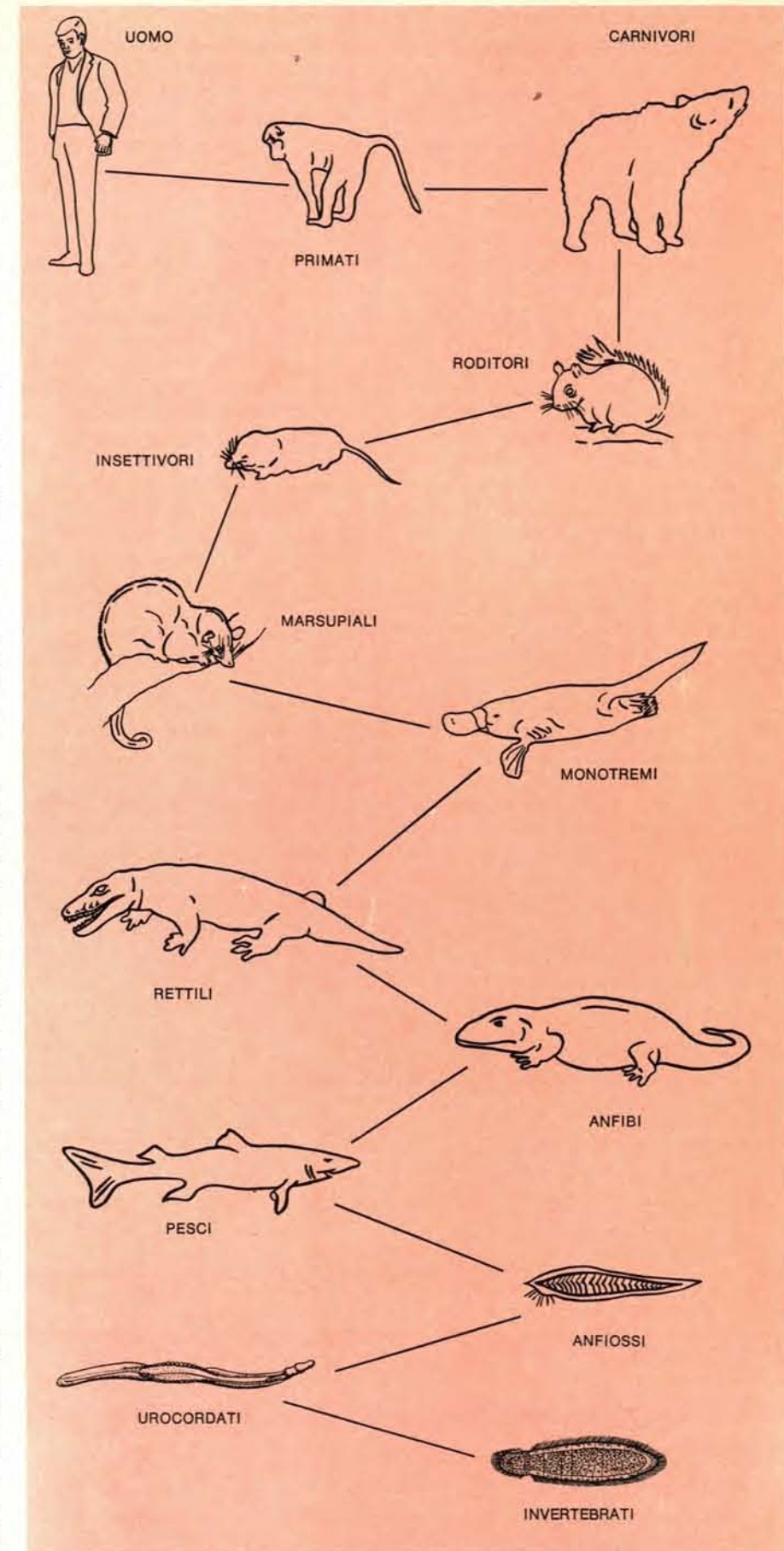
Vi è ancora da considerare l'azione dei neurolettici sull'uomo non normale. Occorre ricordare anzitutto che negli psicotici, a causa della lesione – o dell'inibizione – delle strutture appartenenti al livello geneticamente piú recente, tornano a funzionare le strutture arcaiche disinibite e iperattive: si ha pertanto iperattività motoria (per eccitazione delle strutture motorie arcaiche) e si verificano illusioni o allucinazioni (per eccitazione delle strutture sensoriali arcaiche). Su queste strutture i neurolettici esercitano un'azione inibitoria: eliminano l'agitazione psicomotoria e l'eccitamento maniacale, calmano l'iperattività delle strutture sensoriali in cui sono presenti immagini della vita passata (nelle sindromi amenziali, le allucinazioni sono simili a quelle del sogno) ed esplicano un'azione allucinolitica, senza tuttavia alterare né la motilità normale, né le percezioni.

La teoria dell'evoluzione stratificata consente di mettere in chiaro anche il meccanismo d'azione dell'amfetamina. Di questo farmaco si è molto parlato nel mondo sportivo, a proposito del cosiddetto *doping*. Ora, mi sia lecito dire che tutta la regolamentazione *antidoping*, quando si esca dal campo dell'ippica, è basata su di un errore scientifico. Nell'ippica, il controllo *antidoping* è di vecchia data e ha una sua indiscussa validità: è stato dimostrato, infatti, che il cavallo anche meno dotato diventa capace, sotto l'azione di sostanze amfetaminiche, di battere i piú grandi campioni.

Da qualche tempo, peraltro, si è esteso il controllo *antidoping* a talune competizioni sportive umane. Ora, la azione dell'amfetamina nell'uomo è semplicemente psicotonica: pertanto è fuori di luogo parlare di « droga ». La amfetamina ha affinità con strutture preposte alla motilità, strutture che sono attive nei mammiferi inferiori (e perciò nei cavalli): di qui le grandi corse di « brocchi » ai quali sia stata somministrata una sostanza amfetaminica. Viceversa, nell'uomo l'amfetamina agisce soltanto su taluni centri della sostanza reticolare, che sono centri arcaici integrati, come il centro della veglia, che attiva la corteccia cerebrale; stimolando questo centro l'amfetamina fa dileguare la sensazione di sonno e fa sí che la fatica non venga sentita. In altre parole, sotto l'azione del farmaco un atleta giunge a dare il massimo delle proprie possibilità, ma nulla piú di questo: egli non riuscirà mai a superare un campione, anche se il campione non abbia fatto uso di amfetamina. Dunque

La legge di Haeckel afferma che l'ontogenesi, cioè lo sviluppo dell'individuo, ricapitola la filogenesi (sviluppo della specie): qui ne è rappresentato un esempio attraverso la raffigurazione di quattro embrioni di vertebrati. Da sinistra a destra: elasmobranco (pesce), urodelo (anfibio), uccello, uomo. Negli embrioni dei vertebrati aericoli, ossia in una loro fase ontogenetica, sono presenti le fessure faringee (*b*) necessarie per la respirazione acquatica e proprie della fase filogenetica rappresentata dai pesci.

L'albero genealogico dell'uomo che illustra il trasformismo darwiniano. L'illustrazione è tratta da: Camille Flammarion, *Il mondo prima della creazione dell'uomo*, traduzione italiana di Diego Sant'Ambrogio, Casa editrice Sonzogno, Milano, 1886.

se l'uso di sostanze amfetaminiche può risultare pericoloso per la salute di un atleta, non costituisce tuttavia una frode sportiva (altrimenti, si potrebbe classificare come frode sportiva anche l'uso del caffè): a un atleta modesto l'amfetamina potrà solo consentire, come ho detto, di esprimere il massimo delle sue (modeste) possibilità. Perciò il controllo *antidoping*, prezioso ai fini igienico-sanitari, non ha alcuna ragione di essere sotto il profilo sportivo.

E veniamo alla talidomide. La talidomide, ripetiamo, negli animali da laboratorio (roditori e carnivori) non esercita praticamente alcuna azione; nell'uomo adulto si limita a favorire il sonno, senza alterare le facoltà psichiche. Secondo la mia ipotesi, la talidomide ha affinità con un livello intermedio tra quello attivo nei carnivori e quello attivo nell'uomo; probabilmente un livello attivo nei primati non umani, il livello 12 della classificazione da me proposta. Questa classificazione comprende ventun livelli ontofilogenetici. I primi sei livelli, che nell'ontogenesi maturano nella vita intrauterina, sono, dal punto di vista filogenetico, quelli attivi nei progenitori piú antichi dell'uomo, vissuti nell'èra arcaica, nell'èra primaria e all'inizio dell'èra secondaria. Questi livelli hanno come obiettivo *la sopravvivenza*. I sette livelli successivi hanno come obiettivo *il piacere*: essi comprendono nell'ontogenesi un periodo di circa quattordici mesi, mentre nella filogenesi corrispondono agli stadi dei mammiferi non umani, dal periodo giurassico al Miocene. Gli ultimi sette livelli, che hanno come obiettivo *il bene*, maturano, nell'ontogenesi, dal quattordicesimo mese alla pubertà; nella filogenesi, sono i livelli attivi negli stadi dell'evoluzione che vanno dal Pliocene al nostro tempo (quanto piú recente è uno stadio nella filogenesi, tanto piú lungo è il rispettivo stadio nell'ontogenesi). Finalmente vi è quello che io ho chiamato il livello *extra*, al quale appartiene l'individuo che ha superato le fasi evolutive fissate dall'eredità biologica.

Gli animali da laboratorio appartengono a una fase evolutiva in cui le strutture del livello 12 non sono ancora comparse: ecco perché in questi animali gli effetti della talidomide sono irrilevanti (e sono dovuti, evidentemente, alla modesta affinità del farmaco con i livelli inferiori al livello 12). Quanto all'uomo, la talidomide favorisce il sonno (probabilmente a causa di un'azione inibitrice esercitata dal farmaco sul centro di vigilanza, che è un centro integrato), ma non provoca altri effetti, in quanto nell'uomo le strut-

Secondo la teoria dell'evoluzione stratificata la linea filogenetica che ha portato all'uomo sarebbe quella schematizzata nella figura: dagli invertebrati attraverso gli urocordati e gli anfiossi si arriva ai pesci. Da qui agli anfibi, ai rettili, ai monotremi, quindi ai marsupiali e agli insettivori. Infine ai roditori e carnivori dai quali derivano i primati e l'uomo.

ture appartenenti al livello 12 sono strutture superate dall'evoluzione, ossia strutture che hanno « consegnato » la loro funzione e sono diventate inattive. Ma se una gestante ingerisce la talidomide nel periodo in cui si organizzano nel feto le strutture del livello 12, ecco che il farmaco esplica un'azione teratogena.

Per chiarire gli effetti provocati nel feto da una sostanza ingerita dalla gestante, sarà opportuno ricordare i risultati ottenuti da S. Tusques. Questo autore, somministrando a rane gravide lo ormone tiroideo, constatò che nei nascituri aveva luogo un notevole sviluppo dei lobi olfattivi, del cervelletto, del mesencefalo e della parte ventrale del midollo (in corrispondenza dei motoneuroni degli arti). Parallelamente, Tusques accertò che la somministrazione di ormone tiroideo influenza la formazione di quelle parti del soma che sono innervate dalle strutture nervose con le quali il farmaco ha affinità: ne viene facilitato tanto lo sviluppo degli organi olfattivi quanto lo sviluppo degli arti. Gli esperimenti di Tusques consentono dunque di affermare l'esistenza di una stretta correlazione fra l'azione che un farmaco esercita su talune strutture nervose nella vita embrionale e l'azione che esso esercita su quelle parti del soma innervate dalle stesse strutture.

Prima di tornare all'azione teratogena della talidomide, occorre accennare a quanto si è verificato nel processo evolutivo dopo il raggiungimento dello stadio dei carnivori. Durante l'èra eocenica, alcuni discendenti dei piccoli mammiferi vissuti nel Cretacico divennero belve feroci e sanguinarie. La lotta per l'esistenza si fece sempre piú terribile. Gran parte dei mammiferi inferiori vennero sterminati; altri scamparono alla morte rifugiandosi in zone inaccessibili per i loro nemici. Anche l'antenato dell'uomo cercò riparo nelle foreste, menando vita arboricola (esso scenderà dagli alberi nel Pliocene, quando cioè avrà saputo fabbricarsi gli strumenti con i quali combattere le belve), mentre gli altri mammiferi ritornarono a vivere nell'acqua. L'evoluzione, pertanto, avrebbe preso le seguenti vie: dai carnivori primitivi del Terziario superiore ebbero origine: 1) i carnivori attuali; 2) i mammiferi acquatici; 3) i primati.

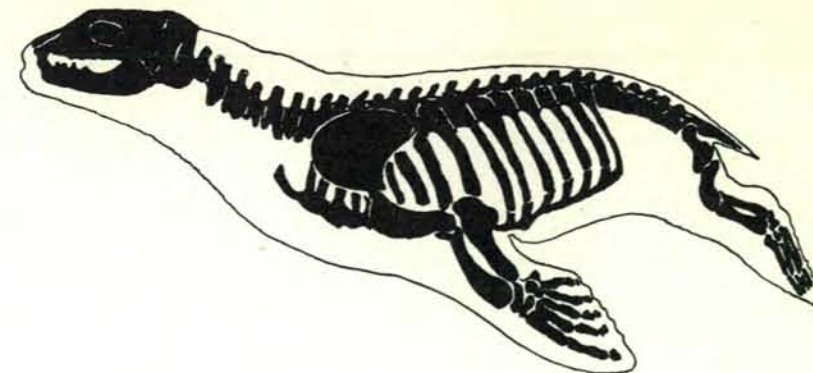
Va anche ricordato che esiste una notevole rassomiglianza tra alcuni ordini di mammiferi acquatici (specie le otarie e le foche) e i carnivori terrestri: la differenza principale consiste nella rudimentazione degli arti, che nei mammiferi acquatici ebbe luogo perché in acqua gli arti erano diventati inutili. In questo stadio dell'evoluzione, pertanto, si assiste ad una modificazione organizzativa degli arti: cessando di avere importanza – o conservandone una infinitamente minore – quali organi di sostegno, gli arti si trasformano: nei primati diventano organi prensili, nei mammiferi acquatici vanno incontro alla rudimentazione.

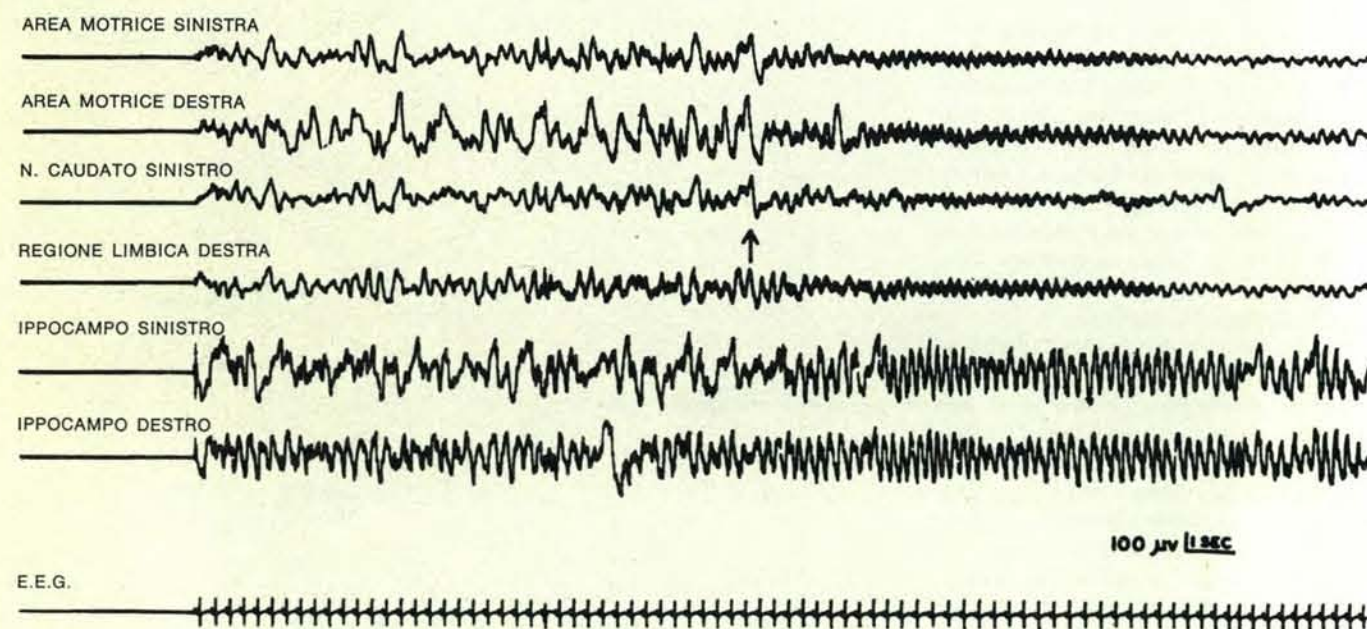
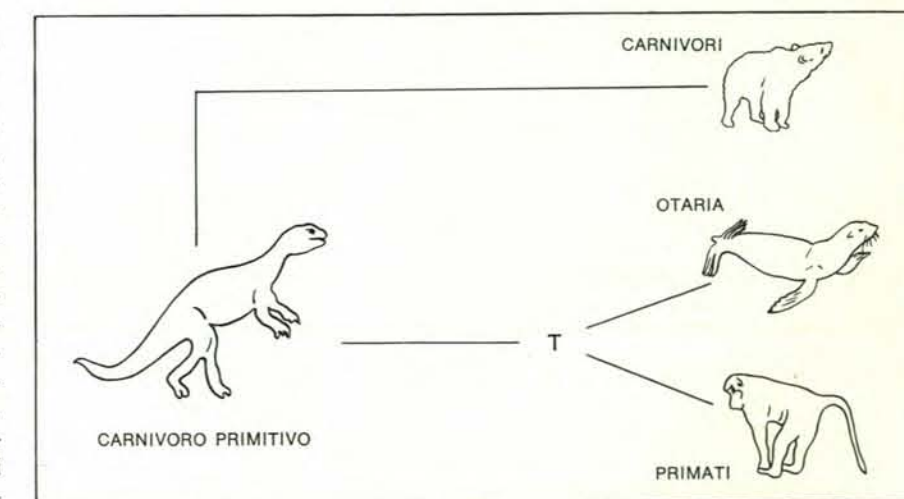
Ora, nella mia ipotesi, la talidomide, inibendo le strutture nervose (e quindi le corrispondenti parti somatiche) che permettono la trasformazione degli arti in organi prensili – la comparsa di polinevriti (a carattere degenerativo) in alcuni soggetti conferma che l'azione della talidomide riguarda determinate strutture, alle quali è legata la motilità degli arti – provoca nell'embrione una deviazione verso una via collaterale dell'evoluzione. (Tale deviazione è limitata alle strutture del livello 12, con le quali la talidomide possiede affinità; pertanto l'organizzazione, nel feto, delle strutture appartenenti ai livelli superiori non viene in alcun modo alterata in seguito alla somministrazione del farmaco alla gestante.)

È questa la via seguita, appunto, dai mammiferi acquatici, che è la rudimentazione degli arti. Di qui la focomelia. In conclusione, il punto di attacco della talidomide potrebbe essere il seguente: la talidomide attacca le sostanze che organizzano gli arti in organi prensili; cosí l'embrione, anziché procedere regolarmente nella direzione che dai carnivori primitivi porta ai primati, subisce una deviazione – per quando riguarda gli arti – verso i mammiferi acquatici.

Ecco dunque il motivo per cui, ancor prima che Schmidt effettuasse gli esperimenti con la talidomide sulle scimmie, io avevo previsto che anche in quegli animali il farmaco avrebbe esercitato, come nell'uomo, una azione teratogena. Mi sia consentito citare quanto scrivevo, quattro anni or sono, nel mio lavoro: « è nostra convinzione, pertanto, che, se gli esperimenti relativi alla talidomide fossero stati effettuati sulle scimmie, anziché sui roditori, l'effetto teratogeno del farmaco sarebbe risultato evidente prima che il farmaco stesso venisse introdotto nella clinica umana ».

I successivi esperimenti sulle scimmie hanno quindi confermato la validità della mia ipotesi. D'altra parte, non è giusto accusare di negligenza criminale i « padri » della talidomide, poiché le conoscenze fin qui acquisite dalla scienza non prescrivono imperativamente che, nei casi in cui un farmaco interessi strutture nervose correlate con le funzioni piú elevate – come le funzioni psichiche – gli esperimenti vengano condotti sulle scimmie, e non su altri animali. Viceversa, un principio basilare al quale ci si dovrà attenere in futuro, se la validità della teoria dell'evoluzione stratificata sarà verificata e comprovata, è appunto il seguente: quanto piú simile è l'organizzazione dell'encefalo di un animale all'organizzazione dell'encefalo umano, tanto maggiori sono le probabilità che i dati ottenuti siano estensibili all'uomo.

Ho avuto modo di esprimere da tempo queste mie conclusioni ai rappresentanti delle maggiori Case farmaceutiche. In proposito, vorrei citare un caso particolare. Una Casa farmaceutica ha sintetizzato un timolettico (sostanza antidepressiva), molto efficace nella psicosi maniaco-depressiva. Ora, uno dei ricercatori ha constatato, e descritto, effetti teratogeni nei nati da ratti, alle cui madri era stato somministrato questo farmaco. Ciò è stato sufficiente ad evocare il disastro della talidomide, e la Casa produttrice del timolettico è giunta a sconsigliare ufficialmente l'uso del farmaco alle gestanti. Ora, io sono convinto che non vi è alcun rapporto tra questo caso e il caso della talidomide. Anzitutto, l'azione teratogena di una sostanza è legata a proprietà inibitorie, mentre i timolettici sono sostanze stimolanti: ciò fa pensare a una coincidenza, per quanto riguarda i dati riscontrati dal ricercatore; tanto piú che nessun altro ha constatato effetti del genere. Ma se anche tali effetti fossero stati confermati, essi sarebbero privi di qualsiasi valore per l'uomo, a meno che non si ripetessero nei primati. Avendo io chiesto ai ricercatori della Casa perché non decidessero di sperimentare il timolettico sulle scimmie (possibilmente antropomorfe), esperimenti che sarebbero risolutivi, mi è stato risposto che il loro costo sarebbe cosí elevato, che la Casa preferisce sconsigliare l'uso del farmaco alle gestanti, piuttosto che affrontare l'onere di esperimenti sulle scimmie: in tal modo, le gestanti che possono averne bisogno restano escluse dai benefici del timolettico. Ora, è noto che l'aspirina esercita nei ratti un'azione teratogena: se l'esperienza di tanti anni non avesse già dimostrato che l'aspirina è innocua per le gestanti, oggi, dopo i fatti della talidomide, l'aspirina sarebbe sconsigliata alle donne incinte!

Per concludere, dobbiamo augurarci che, accanto alle tragiche conseguenze provocate dall'uso della talidomide da parte di donne incinte, ve ne sia almeno una positiva: che in futuro si evitino disastri del genere. Ciò non si potrà ottenere, peraltro, se non si giungerà a mettere in chiaro la causa dell'azione teratogena esercitata dalla talidomide; altrimenti si procederà alla cieca, rinunziando magari a sfruttare le qualità terapeutiche di una sostanza per timore di possibili effetti teratogeni. Vorrei pertanto richiamare l'attenzione degli studiosi e dei ricercatori sull'ipotesi qui prospettata: che, cioè, la talidomide – se ingerita in un particolare periodo della gravidanza – può determinare nell'embrione umano una deviazione, sia pure limitata, del corso dell'ontogenesi. Piú generalmente, vorrei richiamare l'attenzione sull'importanza da attribuire all'evoluzione cerebrale, non solo ai fini della psicofarmacologia, ma ai fini di tutte le scienze che hanno per oggetto fenomeni legati all'attività del sistema nervoso (ivi compresa la neurochirurgia). E, a questo proposito, desidero sottolineare ancora una volta la mia convinzione che è impossibile approfondire compiutamente la conoscenza dell'evoluzione cerebrale se non si pone a confronto l'ontogenesi con la filogenesi, ossia se non si pone a confronto quanto avviene nell'evoluzione individuale con quanto è avvenuto nell'evoluzione della specie.

AREA MOTRICE SINISTRA
AREA MOTRICE DESTRA
N. CAUDATO SINISTRO
REGIONE LIMBICA DESTRA
IPPOCAMPO SINISTRO
IPPOCAMPO DESTRO
100 µv 1 sec
E.E.G.

Gli studi sull'uomo hanno dimostrato che l'attività elettrica cerebrale rimane invariata, o subisce modificazioni insignificanti, dopo iniezione indovenosa di amfetamina. Non cosí, invece, nei mammiferi subprimati: nel coniglio, per esempio, l'amfetamina provoca una accelerazione di tutti i ritmi (*si veda la freccia nel tracciato*) espressione di attivazione della corteccia cerebrale.

Notevole è la rassomiglianza tra alcuni mammiferi acquatici come l'otaria e i carnivori terrestri: la differenza è nella rudimentazione degli arti che nei mammiferi acquatici ebbe luogo perché in acqua, dove questi animali erano ritornati, gli arti erano diventati inutili.

La talidomide inibendo le strutture nervose (e, quindi, le corrispondenti parti somatiche) che permettono la trasformazione degli arti in organi prensili provoca nell'embrione una deviazione verso una via collaterale dell'evoluzione. La figura schematizza il punto d'attacco della talidomide nell'evoluzione filogenetica: T rappresenta il « momento » evolutivo in cui da un carnivoro preistorico evolvono sia la foca sia la scimmia.

L'aborto

L'esperienza di quei paesi in cui l'aborto provocato è legale fornisce i primi dati attendibili su questo metodo di controllo della fertilità, che è vecchio quanto il mondo, il piú diffuso e il piú clandestino.

di Christopher Tietze e Sarah Lewit

La pratica dell'aborto rientra nelle consuetudini umane sin dai tempi piú remoti della piú antica storia scritta. L'aborto è tuttora il metodo di controllo della fertilità piú diffuso, e piú clandestino, del mondo moderno. In questi ultimi anni parecchie nazioni ne hanno legalizzato la pratica e, pertanto, l'aborto provocato sta uscendo dall'ombra e è diventato argomento di discussioni e di controversie in tutto il mondo. La disputa tocca un'ampia gamma di considerazioni: morali, etiche, mediche, sociali, economiche, legali, politiche e umanitarie. L'esperienza dei paesi che hanno reso legale la pratica dell'aborto sta incominciando ora a fornire una quantità di dati attendibili con cui valutare il pro e il contro della pratica stessa.

Che l'aborto provocato fosse probabilmente abituale fra i popoli preistorici ci viene suggerito dagli studi antropologici di tribù primitive che ancor oggi vivono in regioni isolate del mondo. Le motivazioni sono in gran parte le stesse delle società civili: la malattia o l'estrema giovinezza ovvero l'età avanzata della donna incinta, le difficoltà economiche, il desiderio delle donne non sposate di evitare il disonore, e molte altre considerazioni di carattere sociale e biologico. I metodi impiegati dalle donne delle società primitive per interrompere una gravidanza non desiderata vanno dagli sforzi fisici violenti all'applicazione di calore o di irritanti della pelle su varie parti del corpo o all'introduzione nell'utero di una varietà di strumenti.

Di tecniche per abortire si parla già in alcuni dei piú antichi testi medici che ci siano noti. Un'antica opera cinese, che si dice sia stata scritta 4600 anni fa dall'imperatore Shen Nung, contiene una ricetta per provocare l'aborto mediante l'uso di mercurio. Nell'antica Grecia, Ippocrate raccomandava lo sforzo fisico violento come il metodo migliore, e Aristotele e Platone patrocinavano l'aborto per limitare l'aumento della popolazione e mantenere il benessere economico della società. Con l'avvento della religione ebraica e di quella cristiana, l'aborto cadde sotto la condanna morale. Durante il Rinascimento, tuttavia, la disapprovazione popolare dell'aborto si ridusse in modo notevole. Nel diritto consuetudinario inglese l'aborto non veniva considerato un delitto condannabile, a meno che non fosse commesso dopo che la donna aveva avvertito i primi movimenti del feto nel proprio grembo. Anche in questo caso però lo si classificava solo come un semplice atto di cattiva condotta. Soltanto nel 1803 l'aborto divenne in Inghilterra un crimine a termini di legge, mentre negli USA bisognò aspettare fino a circa il 1830. Oggi esso è proibito tuttora nella maggior parte degli stati degli USA, salvo nei casi di grave pericolo di vita per la madre.

Gli argomenti pro e contro l'aborto sono noti a tutti. Le obiezioni di natura morale si basano soprattutto sull'affermazione che la vita umana incomincia con l'unione dell'uovo e dello spermatozoo, sí che la distruzione dell'ovulo fertilizzato è un omicidio, anzi un assassinio. Coloro che si oppongono all'aborto sostengono che esso apre la porta all'abbrutimento della società, incoraggiando le « morti per pietà », l'infanticidio, l'uso delle camere a gas e altre violazioni della santità della vita. Si afferma che l'aborto mina alle basi la struttura sociale, incoraggiando i rapporti sessuali promiscui, indebolendo i legami familiari e facendo sorgere problemi legali in merito al possesso e alla successione dei beni. In campo medico molti dottori trovano disgustosa tale operazione, avvezzi come sono a ritenersi i guardiani della vita umana, e alcuni credono anche che l'aborto possa provocare sterilità, disturbi mestruali, gravidanze ectopiche (quelle cioè in cui l'ovulo si impianta nelle tube uterine), aborti spontanei, parti anormali o sentimenti di colpa che potrebbero portare a sintomi neurotici o persino psicotici.

I difensori dell'aborto sostengono che l'embrione non diventa un essere umano fin dopo un certo periodo di tempo, fissato variamente da 12 a 28 settimane di gestazione. I loro argomenti per la legalizzazione dell'aborto si basano sulla premessa che la cosa di gran lunga piú importante da tener presente dovrebbe essere la salute (mentale e fisica) della futura madre e dell'eventuale figlio. Se esiste una sola possibilità che il bambino possa nascere deforme a causa di fattori genetici, di malattie (per esempio la rosolia durante i primi tre mesi di gravidanza) o di lesioni (come l'ingestione di talidomide da parte della donna incinta), la donna e il suo medico dovrebbero poter decidere se continuare o no la gravidanza. La nascita e l'allevamento di un bambino deforme o debole di mente può avere effetti profondamente sconvolgenti su una famiglia. Si sostiene inoltre che la sua proscrizione, con la conseguente mancanza di norme e di salvaguardie mediche, ha reso l'aborto un grosso problema di salute pubblica in molti paesi, poiché moltissime donne vengono mutilate o uccise da persone inesperte o irresponsabili. I difensori dell'aborto legalizzato fanno rilevare che le tecniche attuali per abortire, quando l'operazione viene eseguita da un medico competente, sono cosí sicure che il rischio per la vita della donna è molto inferiore al rischio dello stesso parto. Alle obiezioni di natura morale si risponde che gli aborti vengono praticati per lo piú su donne sposate e col consenso dei rispettivi mariti, e che una donna violentata non dovrebbe essere costretta a subire l'ulteriore offesa di mettere al mondo il bambino. A livello filosofico, si sostiene che nessuna donna dovrebbe essere costretta a mettere al mondo un bambino non voluto, e che i diritti impliciti di ogni bambino includono quello di nascere desiderato e amato.

Clinicamente possiamo dividere gli aborti provocati in due grandi categorie: quelli eseguiti clandestinamente da gente incompetente e quelli praticati da medici che si avvalgono di tecniche che rendono minimi i rischi per la salute. I metodi impiegati dai profani sono in genere o inefficaci o potenzialmente pericolosi per la donna incinta. Nel corso dei secoli sono stati provati molti tipi di medicazioni interne, che vanno dalla tisana ottenuta dai fiori della *Protea mellifera*, usata dalle tribú analfabete, fino al fosforo bianco e gli ormoni; quasi sempre questi trattamenti non riescono a provocare l'aborto. E non sembra molto probabile che in un futuro abbastanza prossimo

---

In Italia la cronaca registra ogni anno da dieci a venti casi di denunce per attentati alla maternità. Ma il numero di questi reati oscilla, secondo le stime, da 500 000 a 3 milioni ogni anno. Statistiche esatte sono impossibili non solo perché l'aborto è quasi sempre clandestino, ma perché non si è sempre in grado di accertare che sia stato veramente consumato. In caso positivo, è poi difficile distinguere fra aborto procurato e aborto spontaneo.

Il ritardo del flusso mestruale in molti casi è motivo di apprensione e di sgomento, che si cerca di vincere col ricorso a specialità medicinali che dovrebbero provocare l'evento. In Italia, una sola casa farmaceutica denuncia la vendita di oltre un milione di flaconi all'anno di una sua specialità, che si dice abbia efficacia solo se il ritardo mestruale non ha superato i due-tre giorni. È materialmente impossibile una volta comparse le mestruazioni accertare se ciò sia dovuto all'azione abortiva del farmaco oppure a un felice ripensamento della natura. Comunque sia, l'intenzione abortiva è manifesta. Se essa fosse rubricata ai fini statistici, come lo è ai fini religiosi, la cifra di 3 milioni di aborti ipotizzata al 53° Congresso di ostetricia e ginecologia, tenutosi a Bologna l'ottobre scorso, verrebbe con tutta probabilità largamente superata.

Inoltre, quanti degli aborti denunciati come « spontanei », e in buona fede considerati tali, sopraggiunti al 5° o 6° mese di gravidanza, sono conseguenza di pratiche abortive che non avevano dato il frutto sperato e che erano state messe in atto nei primissimi mesi di gravidanza o alcune settimane prima? In casi di questo genere, diventa un compito ben arduo distinguere fra aborto « spontaneo », e perciò non soggetto ai rigori della legge italiana, e l'aborto cosiddetto « criminoso » perché volontario.

Secondo l'AIED, l'associazione italiana per l'educazione demografica, che opera da quindici anni per prevenire, con la diffusione di mezzi antifecondativi sicuri e pratici, nascite illegittime, infanticidi, e soprattutto aborti clandestini, il numero delle interruzioni di gravidanza, ogni anno, in Italia sarebbe attorno al milione. Questo significa che per ogni bambino che nasce c'è un bambino a cui è impedito di venire al mondo.

La stima dell'AIED, che trova conforto nelle statistiche demografiche, in sondaggi di opinione presso medici e in inchieste pluriennali su un campione ristretto ma significativo, deve però ritenersi una stima prudente.

Il tasso di natalità in Italia si è dimezzato negli ultimi sessanta anni: il fascismo – malgrado le ripetute dichiarazioni propagandistiche, che tendevano ad accreditare una realtà diversa – con i suoi incentivi a favore della prolificità non è riuscito ad arrestare la parabola discendente della natalità e oggi l'Italia vanta uno dei piú bassi tassi di natalità fra i paesi europei. È escluso che ciò sia merito di mezzi antifecondativi, come la pillola (si calcola che soltanto 25 000 donne l'abbiano usata nel 1968) o il pessario o gli ovuli vaginali o gli IUD (*Intra-Uterine Devices*) che sono pressoché sconosciuti. È da escludere un'eclisse dell'attività sessuale che, fra gli adolescenti e i giovani, è forse piú intensa di quanto non fosse agli albori di questo secolo. È altrettanto improbabile un'accentuata *impotentia generandi* nei due sessi. La caduta del tasso di natalità deve essere perciò ritenuta come la conseguenza dell'incremento degli aborti.

La società italiana accetta l'aborto con crescente tranquillità, nei ceti piú diseredati (per esempio il sottoproletariato delle borgate romane) e in quelli piú abbienti.

Non ci sono neppure differenze geografiche: non si può dire che le resistenze morali all'aborto siano piú forti nel Sud che nel Nord, in provincia che in città. Se una differenza c'è, questa consiste nelle difficoltà di dare concreta attuazione alla volontà abortiva: minore difficoltà al Nord e nelle città rispetto al Sud e alla campagna. Inoltre le resistenze all'aborto sono forti ancora nel ceto medio, fra i cattolici praticanti. Sembra però che non sia azzardato dire che oggi gli italiani accetterebbero con molta tranquillità e con un certo sollievo una normativa che legalizzasse l'aborto e lo rendesse accessibile a tutti, come è successo in Inghilterra, in URSS e in altri paesi.

L'aborto è anche un fatto economico. Le tariffe variano da 10 000 lire a 500 000 lire. Di 10 000-20 000 lire si accontentano le « praticone » che operano nei paesi e nelle « bidonvilles » cittadine, con mezzi rudimentali e crudeli, come il ferro da calza che talvolta perfora l'utero e provoca strazianti agonie. 80 000-100 000 lire costano gli interventi in ambulatori medici privati: 50 000 lire in piú, se con anestesista. Da 200 000 a 500 000 lire invece l'intervento in clinica. Si calcola in 50 miliardi di lire all'anno il giro di affari sulle interruzioni di gravidanze, sottratti, fra l'altro, ad ogni indagine fiscale.

Nomi ed indirizzi di coloro che praticano aborti sono largamente noti sia all'autorità giudiziaria sia agli organi di polizia, i quali, con molto buon senso, intervengono solo su denuncia precisa, quando la donna muore in seguito all'aborto procurato, oppure quando l'attività dei « fabbricanti d'angeli » dà troppo scandalo.

Se è vero che l'interruzione di gravidanza, praticata da persone esperte con le necessarie precauzioni igienico-sanitarie, presenta meno rischi di un'operazione di appendicite, è anche vero che in Italia sono almeno 20 000 le donne che ogni anno muoiono per postumi di interruzioni dolose di gravidanza.

Di fatto paese degli aborti, l'Italia legale impedisce che si faccia propaganda per prevenire le nascite e ridurre il ricorso all'aborto. I cattolici piú sensibili e provveduti sono favorevoli a *tutti* i metodi antifecondativi, proprio per contenere il fenomeno dell'aborto, sul quale la censura religiosa, non solo cattolica, è assoluta. Ma, nonostante che i partiti laici abbiano presentato in tre legislature un disegno di legge per l'abrogazione dell'art. 553 C.P., il veto alla propaganda antidemografica rimane.

Il parere degli americani sull'aborto si è fatto piú tollerante negli ultimi anni. Il grafico mostra la percentuale delle risposte affermative alla domanda se fosse giusto interrompere la gravidanza per le ragioni prospettate dal questionario dando i risultati per il 1965 e il 1967 e, limitatamente al 1967, il parere dei cattolici (*tinta chiara*), dei protestanti (*piú scura*) e degli ebrei (*la piú scura*). L'aborto viene approvato piú come misura di emergenza che come metodo elettivo di regolazione delle nascite.

si arrivi a produrre « pillole abortive » di sicuro affidamento e non nocive alla salute. La tecnica a cui piú comunemente si ricorre negli Stati Uniti da parte dei non medici è l'inserimento di un corpo estraneo, in genere un tubo di gomma, nel collo dell'utero. Questo sistema provoca contrazioni, emorragia e, alla fine, l'espulsione del feto. Può succedere però che una parte della placenta rimanga nell'utero, e allora l'emorragia non si ferma e la paziente è costretta ad andare da un medico o all'ospedale per l'asportazione, mediante raschiamento, del frammento placentare. In ogni caso, l'operazione non eseguita da un medico comporta in genere un grosso rischio di infezione, a causa della mancanza di precauzioni asettiche.

Tra i medici, la tecnica preferita per l'aborto precoce (vale a dire durante le prime dodici settimane di gravidanza) è quella della dilatazione e del raschiamento o « curettage » (detta piú semplicemente « D e C »). Una serie di dilatatori metallici, ognuno leggermente piú grosso del precedente, viene inserita nel collo dell'utero allo scopo di allargarlo: quando il passaggio è stato aperto a sufficienza, il contenuto dell'utero viene raschiato via con uno speciale strumento. Il tutto richiede in genere solo pochi minuti. In questi ultimi anni è diventato popolare, specie nell'Europa orientale, un particolare metodo: dopo la dilatazione del collo uterino, un tubo di metallo, di vetro o di plastica, che è fornito di un'apertura laterale vicino all'estremità, viene inserito nell'utero e mosso tutt'intorno al fine di staccare l'embrione dalla parete uterina; i frammenti vengono quindi risucchiati per mezzo di una pompa di aspirazione collegata al tubo. Questa tecnica è piú facile, piú rapida e meno traumatica del metodo « D e C ».

Rari sono i casi in cui l'aborto viene praticato dopo la dodicesima settimana di gravidanza, anche nei paesi in cui esso è legalizzato. L'aborto tardivo potrebbe però essere necessario per salvare la vita della gestante. In questo caso il metodo impiegato prevede la rimozione del feto mediante un'operazione simile al taglio cesareo, oppure l'iniezione nell'utero di una soluzione molto concentrata di sale o di zucchero, la quale mette fine allo sviluppo del feto e ne provoca l'espulsione nello spazio di pochi giorni.

In quasi due terzi del mondo (in termini di popolazione) l'aborto è proibito completamente o ammesso solo per ragioni mediche, rigorosamente definite. Gli Stati Uniti sono uno dei paesi restrittivi. In 43 stati l'aborto è permesso solo se la gravidanza minaccia la vita della madre. Gli altri sette stati e il Distretto di Columbia sono un po' piú tolleranti. Nel 1967 e 1968 cinque stati – California, Colorado, Georgia, Maryland e Carolina settentrionale – hanno adottato in materia di aborto leggi piú liberali basate sul codice modello raccomandato dall'American Law Institute. Questo codice permette l'aborto nei casi in cui esista il fondato pericolo che la salute fisica o mentale della madre sia seriamente minacciata, o che il bambino possa nascere con un grave difetto fisico o mentale, o ancora quando la gravidanza è il risultato di una violenza o di un incesto.

La classe medica degli Stati Uniti ha dato in genere un'interpretazione conservatrice delle leggi sull'aborto. I grandi ospedali sottopongono per lo piú ogni domanda di aborto a un'attenta revisione da parte di un consiglio di medici. Si calcola che il numero di aborti terapeutici praticati negli ospedali americani dal 1953 a tutto il 1965 sia stato di solo 8000 circa l'anno, cioè circa due aborti per ogni 1000 nascite. Circa due aborti su cinque furono praticati per ragioni psichiatriche, circa un quarto perché la madre aveva avuto la rosolia durante i primi tre mesi di gravidanza, e il resto per ragioni varie di carattere sanitario. A New York il quoziente di abortività ha oscillato nel 1960-1962 dal 3,9 per ogni mille nascite negli ospedali privati a solo lo 0,1 per 1000 negli ospedali municipali, indicando con ciò che l'aborto legale era meno facilmente alla portata delle famiglie a reddito basso che non a quelle con redditi piú elevati. Vi furono anche notevoli differenze dal punto di vista etnico: la percentuale di aborti terapeutici per 1000 parti fu di 2,6 per le donne bianche, dello 0,5 per quelle negre e dello 0,1 per le portoricane.

Per quanto riguarda il numero degli aborti illegali praticati negli Stati Uniti è estremamente difficile avere una valutazione attendibile. Nel 1955 si tenne a Arden House una conferenza sull'aborto, organizzata dalla Planned Parenthood Federation e dalla New York Academy of Medicine; un comitato della conferenza potè concludere soltanto che il numero degli aborti illegali potrebbe essere pari alla bassa cifra di 200 000 o toccare invece quella di 1 200 000 l'anno. Da allora le cose non sono cambiate, e non si dispone ancora di informazioni su cui basare una valutazione migliore. E non si hanno nemmeno dati attendibili per determinare il numero di morti materne da aborto illegale negli USA. Una trentina d'anni fa si pensava che tali morti potessero ammontare a 5000-10 000 l'anno, ma questa cifra, anche se abbastanza precisa a quel tempo, non può in nessun modo avvicinarsi alla vera cifra odierna. Il numero totale di morti da tutte le cause fra le donne in età feconda non supera negli Stati Uniti la cifra di circa 50 000 l'anno. Il National Center for Health Statistics ha registrato, nel 1965, 235 morti conseguenti ad aborto. La mortalità totale da aborto illegale fu indubbiamente piú alta di questa cifra, ma con tutta probabilità fu inferiore a 1000.

Il compianto Alfred C. Kinsey e i suoi collaboratori, nei loro famosi studi sul comportamento sessuale svolti piú di un decennio fa, trovarono che fra le donne intervistate l'aborto veniva praticato nella maggior parte dei casi (90 per cento circa) da un medico, benché non necessariamente da un medico stimato. Si sospetta che molte donne americane che intendono abortire vadano all'estero, in paesi in cui le leggi non sono cosí restrittive o non vengono applicate. Per le donne meno privilegiate che non possono ricorrere a un medico abile, l'aborto presenta rischi paragonabili agli altri rischi che si possono correre in una gravidanza. Secondo uno studio condotto da Edwin M. Gold sulle morti causate da complicazioni della gravidanza o del parto nella città di New York nel 1960-1962, l'aborto fu causa di morte per il 25 per cento delle donne bianche, per il 49 per cento di quelle non bianche e per il 56 per cento delle portoricane.

Dopo la conferenza di Arden House del 1955, l'aborto è stato oggetto continuo di discussione da parte di molti gruppi professionali e civili. Nel 1967 la House of Delegates dell'American Medical Association ha approvato la legalizzazione delle leggi sull'aborto, e nel 1968 l'American Public Health Association ha raccomandato l'abrogazione di tutte le leggi restrittive.

Anche i sondaggi dell'opinione pubblica hanno rivelato una tendenza favorevole alla liberalizzazione. Nel National Fertility Study condotto nell'autunno del 1965, una notevole maggioranza si dichiarò in favore della legalizzazione dell'aborto per proteggere la salute della madre, e una maggioranza piú piccola approvò l'aborto per impedire la nascita di un bambino anormale o in casi di violenza. D'altra parte, solo il 13 per cento approvò a quel tempo l'aborto per le donne non sposate, e solo l'8 per cento per qualsiasi donna che lo volesse. Due anni dopo, un sondaggio Gallup svolto per conto del Population Council rivelò un « atteggiamento molto piú liberale » da parte del pubblico, con un incremento fino al 28 e al 21 per cento del numero di coloro che approvavano rispettivamente l'aborto non solo per le donne non sposate ma anche per chiunque ne facesse richiesta.

I due terzi del mondo in cui sono in vigore proibizioni analoghe a quelle che si hanno negli USA comprendono paesi dell'emisfero occidentale, dell'Europa meridionale e occidentale, dell'Africa e della maggior parte dell'Asia, a eccezione del Giappone e della Cina. In Inghilterra e nei paesi scandinavi, che costituiscono il 2 per cento della popolazione mondiale, le leggi sono notevolmente piú liberali, in quanto permettono l'aborto in base a un'ampia gamma di indicazioni mediche, eugenetiche e umanitarie. Il Giappone, la Cina e la maggior parte dei paesi dell'Europa orientale, che costituiscono quasi il rimanente terzo della popolazione mondiale, sono andati molto piú avanti sulla via della liberalizzazione, permettendo l'aborto a richiesta della donna incinta o in base a indicazioni sociali di ampia interpretazione.

I paesi con la piú lunga esperienza ininterrotta sono la Svezia e la Danimarca, che incominciarono a liberalizzare le proprie leggi sull'aborto nel decennio 1930-1940, e che con l'andar degli anni hanno diminuito a poco a poco le condizioni necessarie per poter abortire. In entrambi i paesi i criteri per concedere il permesso si basano innanzitutto sui rischi fisici e mentali della madre, tenendo presente « le condizioni nelle quali la donna dovrà vivere ». In realtà, la maggior parte degli aborti sono stati approvati in base a motivazioni di natura psichiatrica. In Svezia, dove un tempo le domande venivano rivolte per lo piú al Consiglio Medico Reale, le operazioni vengono ora eseguite in gran parte su raccomandazione di due medici. In Danimarca un certificato rilasciato dal capo del reparto ostetrico e ginecologico di un ospedale è sufficiente ad autorizzare un aborto in base a considerazioni esclusivamente mediche; in casi implicanti altri fattori la domanda deve essere approvata da un comitato composto da uno psichiatra, da un ginecologo e da un assistente sociale.

Il numero annuo di aborti legali salí in Svezia da 400 nel 1939 a oltre 6300 nel 1951; poi scese di nuovo, per tornare a salire negli ultimi pochi anni, fino a toccare la punta massima di 9600 nel 1967. La percentuale degli aborti rispetto ai nati vivi è stata di cinque per 1000 nel 1939 e di 79 per 1000 nel 1967. L'esperienza della Danimarca è stata parallela a quella della Svezia (*si veda l'illustrazione nella pagina a fronte*). Via via che aumentavano gli aborti legali, diminuiva il tasso di mortalità associato con l'operazione. In Svezia, per esempio, la mortalità scese da 257 per 100 000 aborti legali nel periodo compreso tra il 1946 e il 1948 a 39 per 100 000 nel periodo 1960-1966.

In Inghilterra l'aborto praticato per proteggere la salute della madre è legale sin da una sentenza del 1938, relativa a un famoso caso di violenza in cui il giudice dichiarava che l'aborto era giustificato « se il dottore ritiene... che le probabili conseguenze [della nascita] possano finire col rendere la madre un rottame fisico e mentale ». La classe medica inglese, tuttavia, assunse un atteggiamento conservatore nei confronti di questa sentenza. Nel 1967, dopo un'aspra e lunga lotta, il Parlamento approvò una legge sull'aborto simile al modello scandinavo, il quale veniva persino superato in certi punti importanti. La nuova legge inglese autorizza l'aborto quando due medici certificano che la continuazione della gravidanza comporterebbe dei pericoli per la vita o per la salute fisica o mentale della donna incinta, anche se il pericolo è molto lieve: basta solo che esso sia piú grande di quello dell'aborto che, se eseguito nelle condizioni piú convenienti, presenta minor pericolo per la vita che non la gravidanza e il parto. La legge inglese permette inoltre l'aborto qualora l'eventuale nascita possa mettere in pericolo la salute fisica o mentale di « bambini già esistenti » nella famiglia della gestante, o qualora ci sia il fondato pericolo che il bambino nasca anormale. Dall'entrata in vigore della nuova legge britannica, nell'aprile del 1968, non si hanno informazioni significative in merito ai suoi effetti.

Negli USA l'aborto provocato è in genere illegale, ma in certe circostanze può anche essere eseguito negli ospedali. È piú alla portata della gente abbiente, come risulta da questo grafico che dà la percentuale degli aborti a New York dal 1960 a tutto il 1962.

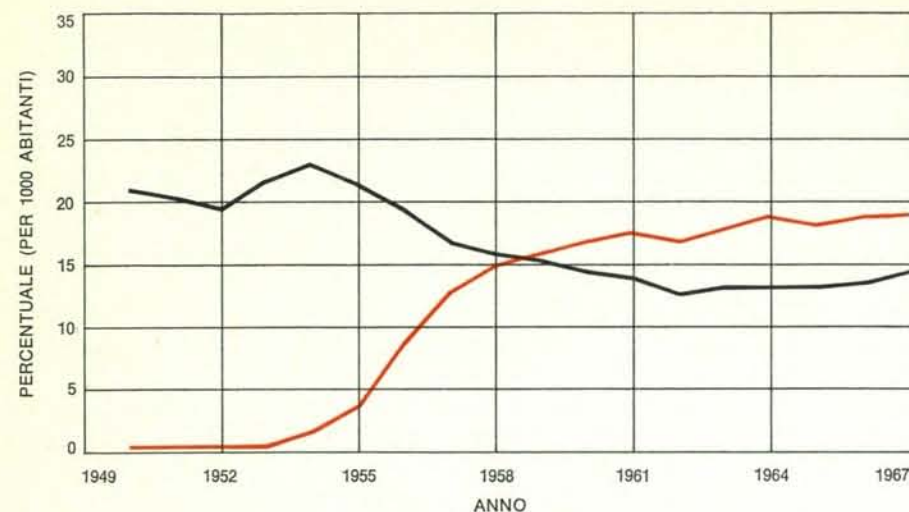
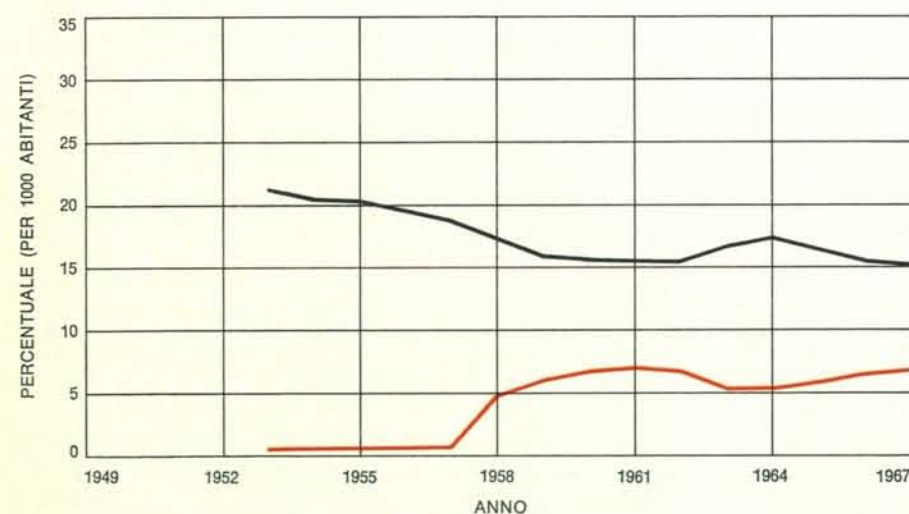
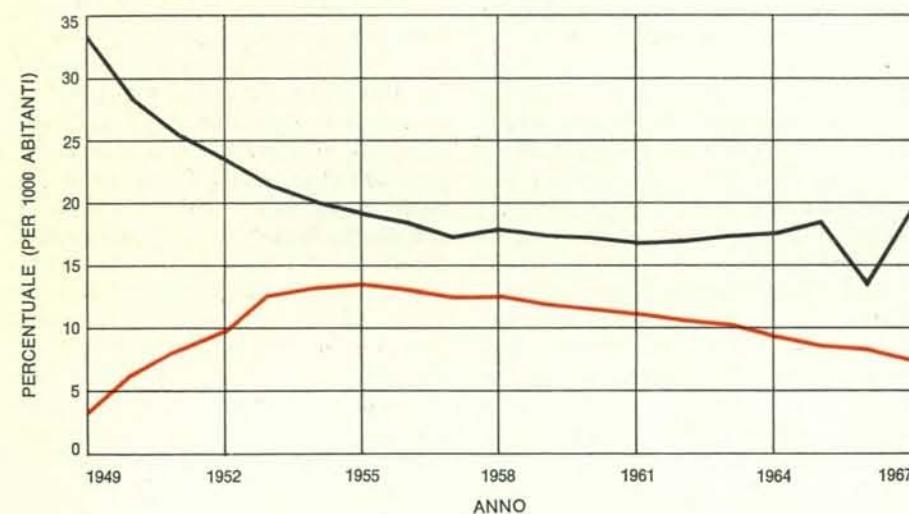
Qual è stata l'esperienza dei paesi con le leggi sull'aborto piú tolleranti, dei paesi cioè dell'Europa orientale, del Giappone e della Cina? Sulla Cina non si hanno informazioni di nessun genere, mentre si hanno dati abbastanza particolareggiati su paesi come l'Ungheria e la Cecoslovacchia, che (insieme con l'URSS) hanno approvato nel decennio 1950-1960 leggi sull'aborto molto liberali, e sul Giappone, che ha una « legge per la protezione eugenetica » sin dal 1948.

In quasi tutti questi paesi le leggi permettono l'aborto in base a indicazioni di carattere sociale interpretate con molta liberalità, se non addirittura su richiesta della gestante. La Polonia, per esempio, registra qualunque « situazione sociale difficile » come ragione accettabile per concedere l'aborto legale. La Cecoslovacchia permette l'aborto in base all'età avanzata della donna, alla presenza di tre o piú bambini nella famiglia, alla perdita o all'invalidità del marito, a una famiglia divisa, a una minaccia al tenore di vita della famiglia quando le principali responsabilità economiche ricadono sulla donna, e a « circostanze difficili », come nel caso di una donna non sposata e di una gravidanza conseguente a violenza. Analogamente, in Jugoslavia l'aborto è ammesso ogniqualvolta la nascita di un bambino creerebbe inevitabilmente alla madre gravi difficoltà di natura personale, familiare o economica. Di norma, nei paesi dell'Europa orientale gli aborti giustificati dal punto di vista medico vengono praticati gratuitamente negli ospedali; negli altri casi la donna paga una parte delle spese.

Non c'è da sorprendersi se, là dove le leggi sull'aborto sono tolleranti, cosí come lo sono nell'Europa orientale, il numero degli aborti legali è salito enormemente. In Ungheria, per esempio, l'incremento è stato cosí grande, che il numero degli aborti legali supera ora quello delle nascite. Nel 1967 si ebbero in quel paese 187 500 aborti legali contro 148 900 nati vivi (*si veda l'illustrazione in alto nella pagina seguente*). L'andamento opposto delle curve, da cui risulta che il numero di aborti è andato aumentando mentre quello delle nascite è andato diminuendo, fa pensare che per molte donne ungheresi l'aborto abbia preso il posto delle pratiche anticoncezionali come mezzo di controllo delle nascite. Analogamente, le statistiche cecoslovacche indicano che l'aborto legalizzato ha ridotto la natalità, benché questa sia tuttora notevolmente superiore all'indice di abortività (*si veda l'illustrazione di mezzo nella pagina seguente*).

I paesi scandinavi hanno continuato a liberalizzare le leggi sull'aborto sin dal decennio 1930-1940. Il grafico mostra la curva dell'indice di abortività in Svezia (*nero*) e Danimarca (*colore*) nel periodo che va dal 1939 al 1967. Le punte massime furono raggiunte negli anni cinquanta, vi fu poi una flessione nei primi anni del decennio successivo, seguita ora da un nuovo aumento.

Il calo del numero delle nascite ha spinto due paesi dell'Europa orientale a rivedere le proprie leggi relative all'aborto. Nell'ottobre del 1966, la Romania ha sostituito la legge che permetteva l'aborto a richiesta con una di carattere restrittivo. L'aborto in base a giustificazioni mediche viene ora permesso solo se la gravidanza minaccia la vita della donna, e le altre indicazioni accettabili sono limitate ai casi in cui la donna ha piú di 45 anni, o ha quattro figli, o rischia di mettere al mondo una creatura deforme, o è rimasta incinta in seguito a violenza o a incesto. Dopo l'approvazione della nuova legge rumena, la natalità salí dal 13,7 (per 1000 abitanti) del quarto trimestre del 1966 al 38,4 del terzo trimestre del 1967. Questo incremento indica che, prima dell'abrogazione della legge piú tollerante, il numero annuo di aborti legali in Romania era stato di almeno 24,7 per ogni 1000 abitanti, cifra notevolmente superiore all'alto tasso del 18,4 per mille registrato in Ungheria nel 1967.

Nel 1968 anche la Bulgaria impose delle restrizioni – un po' meno drastiche – alle possibilità di aborto legale. Le donne con tre o piú figli possono ancora procurarsi l'aborto a richiesta, ma quelle con uno o due figli debbono fare domanda a una commissione, la quale « spiegherà i danni e i pericoli dell'aborto, la necessità di portare a termine la gravidanza e il sostegno finanziario che la famiglia riceverà dopo la nascita di un figlio. In genere, questa commissione farà ogni sforzo per far recedere dal suo intento ogni donna che esprima il desiderio di interrompere la gravidanza. Se però, nonostante tutto, la persona interessata persiste nel chiedere l'interruzione della gravidanza, la commissione darà la propria approvazione in merito ». Le donne senza figli possono abortire solo per ragioni mediche.

In Ungheria, dove l'aborto viene concesso su richiesta, ed è un mezzo normale di controllo delle nascite, il numero degli aborti (*colore*) supera quello delle nascite (*nero*).

In Cecoslovacchia si concede l'aborto per ragioni mediche, sociali ed economiche. Ultimamente la percentuale di aborti è salita (*colore*) quella delle nascite (*nero*) è calata.

In Giappone, dove l'aborto è alla libera portata di tutti, può darsi che la percentuale degli aborti (*colore*) appaia ridotta nelle statistiche artificiosamente. Ciò è indicato dal fatto che essa non aumentò quando la natalità (*nero*) subí una brusca flessione nel 1966.

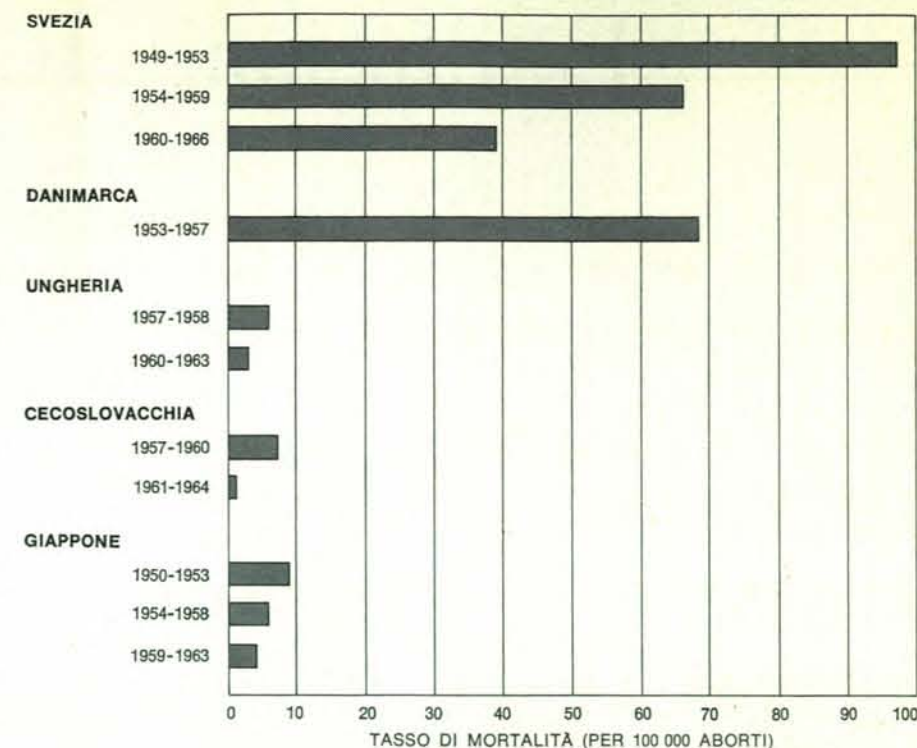
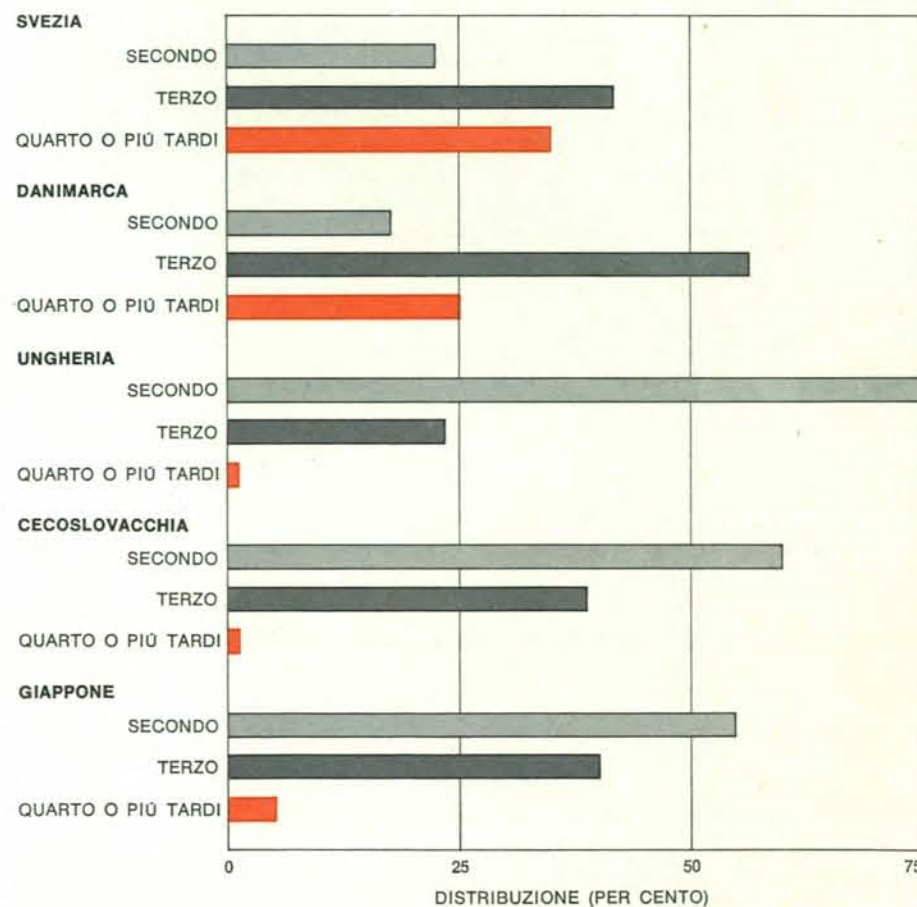
In Giappone la legge per la protezione eugenetica autorizza l'aborto nei casi in cui la gravidanza influisca seriamente sulla salute della donna « sia dal punto di vista medico, sia da quello economico ».

Questa clausola è stata interpretata con tanta elasticità che qualunque donna giapponese può praticamente avere un aborto legale a richiesta. Il numero degli aborti legali registrati in Giappone è salito da 246 000 nel 1949, l'anno successivo all'approvazione della legge, a 1 170 000 nel 1955. Dopo questa data, però, il numero registrato ha subito una flessione e nel 1967 era sceso a 748 000 (*si veda l'illustrazione in basso a sinistra*). Vi sono tuttavia fondate ragioni per credere che le cifre ufficiali siano inferiori di parecchie centinaia di migliaia al vero numero di aborti legali. Si dice infatti che molti medici giapponesi, allo scopo di ridurre al minimo l'imposta sul reddito, non denuncino tutti gli aborti eseguiti. Cosa strana, il numero degli aborti registrati continuò a diminuire invece di aumentare quando nel 1966, l'« Anno del Cavallo Focoso », le nascite calarono a precipizio. (Secondo la tradizione, le bambine nate in tale anno avranno da grandi un brutto carattere, donde la difficoltà di trovar loro un marito). Siccome è dubbio che la brusca flessione delle nascite possa essere attribuita principalmente a un uso piú efficace delle pratiche anticoncezionali da parte di quel settore della società giapponese piú legato alle tradizioni, è lecito supporre che nel 1966 il numero degli aborti in realtà sia cresciuto, e non diminuito.

Un fatto degno di nota a proposito della pratica legalizzata dell'aborto nei paesi socialisti dell'Europa orientale e in Giappone è la mortalità molto bassa. In questi paesi si sono registrate negli ultimi anni solo da uno a quattro morti conseguenti ad aborto per ogni 100 000 aborti legali, una percentuale cioè molto piú bassa di quella dei paesi scandinavi (*si veda l'illustrazione in alto nella pagina accanto*). Una delle ragioni principali di questo piú basso indice di mortalità è che il Giappone e i paesi dell'Europa orientale permettono l'aborto dopo il terzo mese di gravidanza soltanto per ragioni mediche, laddove la Svezia, per esempio, lo permette fino al quinto mese e la Danimarca fino al quarto. Una altra ragione è il piú cattivo stato di salute delle donne che chiedono di abortire nei paesi scandinavi rispetto a quello delle donne che si sottopongono a richiesta all'intervento abortivo nei paesi dell'Europa orientale e in Giappone.

È difficile determinare in quale misura la legalizzazione dell'aborto in questi paesi abbia ridotto l'incidenza degli aborti criminosi, o clandestini. Certo gli aborti illegali non sono stati eliminati completamente, come è dimostrato dal fatto che molte donne si rivolgono tuttora agli ospedali per le cure dopo un aborto illegale. Questa tenace sopravvivenza di aborti illegali può essere forse attribuita al fatto che, per una ragione o per l'altra, le donne a volte vogliono nascondere la propria gravidanza, e quindi evitano la procedura della domanda formale e del permesso ufficiale. Comunque sia, non c'è dubbio che nei paesi che permettono l'aborto a richiesta la pratica dell'aborto illegale ha subito una grande flessione.

Poco si sa sulla situazione nel resto del mondo. Già si è notato come negli USA il numero degli aborti illegali sia calcolato a centinaia di migliaia l'anno. Nella maggior parte dei paesi dell'Europa occidentale si pensa che l'aborto sia piú comune che negli USA, e è stato detto, ma non verificato, che in alcuni di questi paesi il numero degli aborti superi quello delle nascite. L'aborto illegale appare molto frequente nell'America Latina, specialmente fra i poveri delle città. Nel Cile, le cui leggi relative all'aborto sono simili a quelle degli USA, Rolando Armijo e Tegualda Monreal, della Scuola della Sanità Pubblica dell'Università del Cile, hanno concluso, dopo un attento studio delle statistiche anagrafiche, che a Santiago, la capitale del paese, si sono avuti in un solo anno 20 000 aborti illegali contro 77 440 nascite. Una situazione analoga è stata registrata dall'altra parte del mondo. In Corea, che ha una legge restrittiva sull'aborto simile a quelle della maggior parte degli stati degli USA, Sung-bong Hong ha calcolato che a Seul, la capitale, si sono avuti 55 000 aborti illegali – piú della metà delle nascite – che ammontavano a 108 700.

Quello dell'aborto illegale è un problema che, in un modo o nell'altro, ogni nazione deve pur affrontare alla fine. In molti paesi esso è cosí grave che i medici e il governo hanno unito le forze per organizzare programmi di pianificazione della famiglia. C'è da sperare che gli USA, e in particolar modo i corpi legislativi dei vari stati, si avvalgano a questo proposito dell'esperienza che viene ora offerta dai paesi che stanno tentando nuove strade nella ricerca di una soluzione razionale del problema.

Il tasso di mortalità è noto solo dove gli aborti sono legali. È molto piú basso nei tolleranti paesi dell'Europa orientale e in Giappone che non in Svezia e in Danimarca.

I bassi indici di mortalità dell'Europa orientale e del Giappone vengono attribuiti al fatto che in questi paesi si cerca di evitare che gli aborti siano effettuati a gravidanza inoltrata. Nel grafico gli aborti sono distribuiti per mese di gestazione.

Le illusioni ottiche

Perché le figure semplici appaiono talvolta distorte o complesse? Forse perché il sistema visivo è costretto a interpretare un mondo nel quale gli oggetti sono normalmente distorti dalla prospettiva.

di Richard L. Gregory

Una soddisfacente teoria della percezione visiva deve essere in grado di spiegare in che modo fugaci immagini luminose, giungendo alla retina dell'occhio, comunichino la conoscenza degli oggetti esterni. Come il cervello « legga » la realtà dalle immagini oculari è difficile a dirsi poiché gli oggetti sono molto differenti dalle immagini che rappresentano in maniera diretta solo alcune delle loro caratteristiche importanti. In ogni istante, infatti, l'immagine retinica rappresenta il colore di un oggetto e la sua forma da una singola posizione, ma il colore e la forma sono in se stessi dei dati superficiali. Il colore dipende dalla qualità dell'illuminazione e da fattori più elusivi quali il contrasto e l'affaticamento retinico. La forma, come tutti sappiamo, può essere fortemente distorta da illusioni di vario tipo. Dato che, ovviamente, il maggior interesse di chi si serve degli occhi non è certo quello di essere ingannato da illusioni ottiche, si cerca di conoscere in che modo le illusioni stesse si verifichino. Le illusioni potrebbero determinarsi durante il processo di elaborazione delle informazioni che, in condizioni normali, rende il mondo visibile più facile da comprendere? Questa è la proposizione base che cercherò di esaminare.

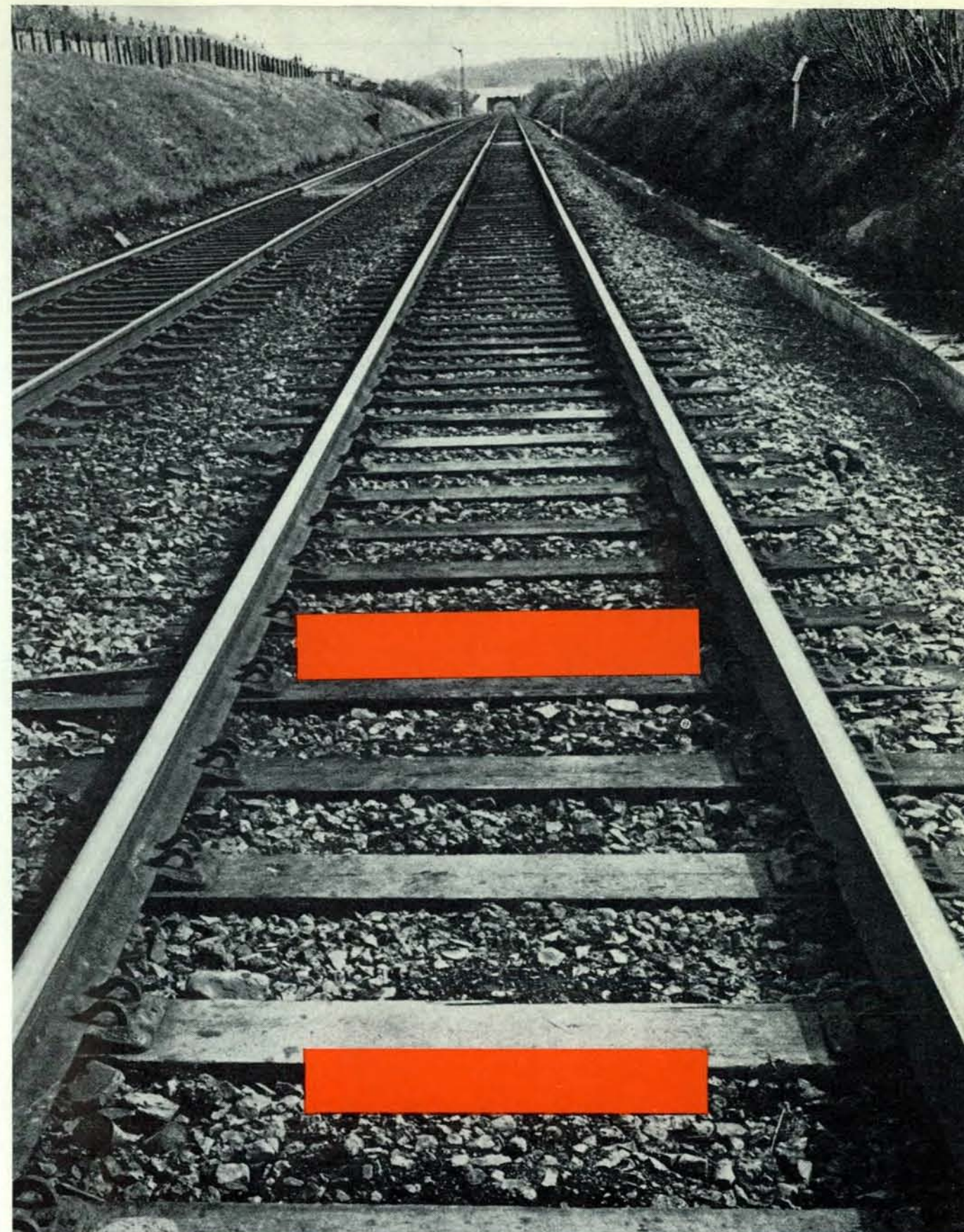
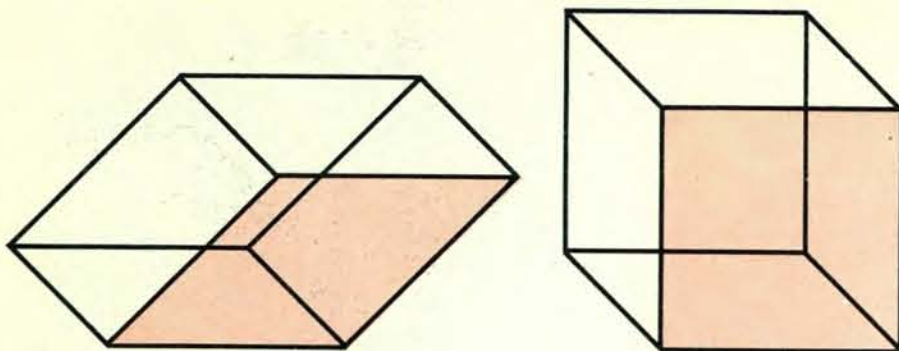
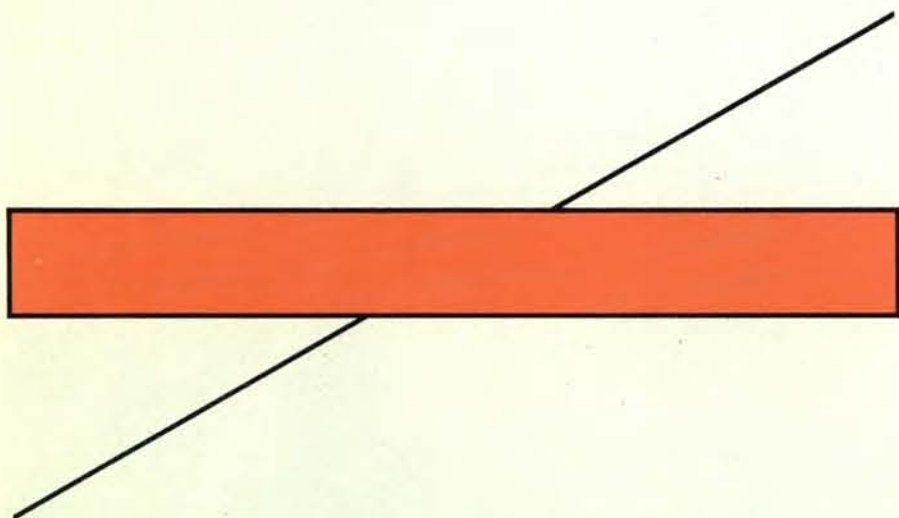
Illusioni di vario genere possono verificarsi in ognuno dei sensi e possono trasferirsi da un senso all'altro. Per esempio, gli oggetti minuscoli sembrano notevolmente più pesanti di oggetti più grandi aventi esattamente lo stesso peso. Lo si può dimostrare facilmente riempiendo di sabbia un piccolo barattolo e mettendone altrettanta in un barattolo più grande. Il recipiente più piccolo potrà sembrare fino a una volta e mezzo più pesante del recipiente più grande che ha esattamente lo stesso peso. È evidente, perciò, che il peso è percepito non soltanto come pressione e sensazione muscolare ma anche in funzione del peso che si presume abbia l'oggetto, in base alle sue dimensioni stimate a vista. Quando la densità non è nota, la visione produce l'illusione del peso. Io ritengo che tutte le illusioni di distorsione sistematica siano fondamentalmente simili all'illusione dimensione-peso.

Benché parecchie illusioni ottiche fossero note agli antichi greci, esse sono state studiate sperimentalmente soltanto da poco più di un secolo. La prima descrizione scientifica moderna si trova in una lettera al fisico scozzese Sir David Brewster del naturalista svizzero L.A. Necker, il quale nel 1832 scrisse che il disegno di un romboide trasparente si rovescia in profondità: talvolta appare direttamente di fronte una faccia e talora l'altra. Necker notò che, sebbene i cambiamenti nella fissità dell'occhio potessero indurre tale mutamento nella percezione, esso si sarebbe potuto manifestare in maniera assolutamente spontanea. Questo celebre effetto è illustrato di solito con un cubo isometrico anziché con la figura originale di Necker (*si veda l'illustrazione in alto a pagina 32*).

Alquanto più tardi W. J. Sinsteden segnalò un effetto parimenti sorprendente che doveva essere ben noto da tempo agli olandesi. Se le pale rotanti di un mulino a vento vengono viste obliquamente o direttamente dal fianco, sembrano invertire spontaneamente direzione quando non vi sono evidenti indizi sulla direzione di rotazione. Questo effetto può essere dimostrato proiettando su uno schermo l'ombra, vista in prospettiva, di una pala che ruota lentamente. In assenza di ogni indizio che segnali la direzione di rotazione, la pala sembrerà invertire spontaneamente la direzione e a volte l'ombra sembrerà anche contrarsi ed espandersi. È importante notare che questi effetti non sono distorsioni percettive dell'immagine retinica; essi sono interpretazioni alternative dell'immagine in termini di oggetti possibili. È come se il cervello prendesse in considerazione ipotesi diverse sull'oggetto che può corrispondere all'immagine oculare. Quando i dati sensori sono insufficienti, sono prese in considerazione diverse ipotesi e il cervello « non riesce a prendere mai » una decisione definitiva.

Le illusioni ottiche più complicate sono le distorsioni sistematiche di grandezza o forma. Queste distorsioni si verificano in parecchie figure molto semplici. La distorsione ha lo stesso andamento e ricorre all'incirca in ugual misura in tutti gli osservatori umani e probabilmente anche in molti animali. Queste distorsioni costituiscono un'importante sfida per gli psicologi, in quanto devono venire spiegate da una teoria soddisfacente della percezione normale e in quanto potrebbero costituire importanti indizi sui processi percettivi fondamentali.

Illusioni di distorsione

La più semplice illusione di distorsione fu anche la prima a essere studiata. Si tratta dell'illusione orizzontale-verticale descritta da Wilhelm Wundt, assistente di Herman von Helmholtz a Heidelberg e considerato il padre della psicologia sperimentale. L'illusione è costituita semplicemente da una linea verticale che sembra più lunga di una linea orizzontale di eguale lunghezza. Wundt attribuí la distorsione all'asimmetria del sistema che muove l'occhio. Per quanto questa spiegazione sia stata invocata molte altre volte, deve essere scartata poiché le distorsioni avvengono sia nelle immagini residue (o con-

L'illusione implicante la prospettiva è in generale costante per tutti gli osservatori umani. I due rettangoli sovrapposti a questa fotografia di binari ferroviari sono esattamente delle stesse dimensioni, eppure il rettangolo superiore sembra notevolmente più grande. L'autore considera questa illusione come un prototipo delle distorsioni visive nelle quali il meccanismo percettivo, che comprende il cervello, tenta di mantenere un'approssimativa costanza dimensionale per oggetti simili posti a distanze differenti. Giacché sappiamo che le traversine in lontananza sono grandi come quelle più vicine, un qualsiasi oggetto messo fra i binari a distanza intermedia (il rettangolo superiore) viene inconsciamente espanso. Infatti, se i rettangoli fossero oggetti reali giacenti fra le rotaie, sapremmo immediatamente che il rettangolo più distante è il più grande.

L'illusione di L.A. Necker. Egli notò che un romboide trasparente (*a sinistra*) si rovescia spontaneamente in profondità. L'area in colore leggero può apparire sia come una superficie esterna sia come una superficie interna di una scatola trasparente. Oggi di solito si usa un cubo (*a destra*), noto come cubo di Necker, invece del romboide.

L'illusione di Poggendorff fu proposta da Johann Poggendorff nel 1860, lo stesso anno in cui Johann Zöllner propose la figura mostrata in copertina. Nella figura di Poggendorff i due segmenti della linea diagonale sembrano non trovarsi sulla stessa retta.

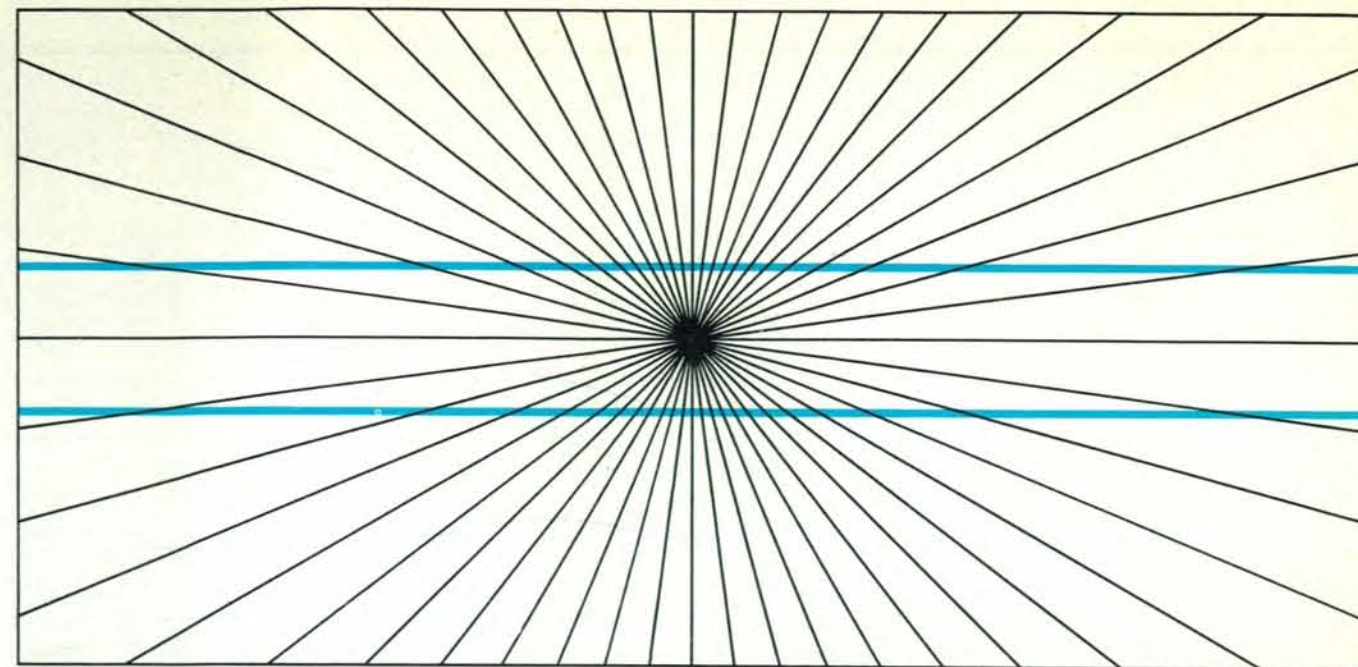
secutive) sulla retina sia nelle immagini normali stabilizzate artificialmente in modo che rimangano ferme sulla retina. Inoltre, le distorsioni possono contemporaneamente verificarsi in diverse direzioni, il che può difficilmente essere attribuito ai movimenti dell'occhio. È anche difficile immaginare in che modo le distorsioni di curvatura possano essere attribuite ai movimenti dell'occhio. L'evidenza suggerisce quindi che le distorsioni non si originino negli occhi ma nel cervello.

L'interesse per le illusioni divenne generale con la pubblicazione di parecchie figure che segnalavano le distorsioni che avrebbero potuto produrre errori nell'impiego degli strumenti ottici. Questi errori costituirono una grande preoccupazione per i fisici e gli astronomi di cent'anni fa, quando la fotografia e gli altri mezzi per eliminare gli errori visivi erano ai primi passi. La prima delle speciali figure di distorsione fu l'illusione pubblicata da Johann Zöllner nel 1860 (*si veda la figura in copertina*). Lo stesso anno Johann Poggendorff pubblicò la sua illusione relativa all'apparente spostamento di una linea (*si veda l'illustrazione qui sopra*). Un anno più tardi Ewald Hering presentò l'illusione, oggi ben nota, nella quale linee parallele appaiono incurvate (*si veda l'illustrazione nella pagina a fronte*); l'illusione opposta fu concepita nel 1896 da Wundt.

La piú famosa di tutte le illusioni di distorsione è probabilmente la freccia a due punte di Franz Müller-Lyer, presentata in 15 variazioni nel 1889 (*si veda l'illustrazione a pagina 34*). Questa figura è cosí semplice e la distorsione cosí evidente che fu immediatamente impiegata come bersaglio principale di ipotesi ed esperimenti. Furono proposte teorie di ogni tipo, e Wundt invocò di nuovo la sua teoria del movimento oculare. Alcuni supposero che le « ali » poste sulle punte della freccia allontanassero l'attenzione dagli estremi della linea centrale provocando in tal modo il suo espandersi o contrarsi. Altri considerarono la distorsione in base a un principio generale per il quale gli angoli acuti tendono a essere sovrastimati e gli ottusi sottovalutati, ma non riuscirono a spiegarne la ragione.

Tutte queste teorie avevano una caratteristica comune: erano dei tentativi per spiegare le distorsioni in termini di forma dello stimolo, senza riferimenti al possibile significato in termini di percezione degli oggetti. Vi fu, tuttavia, una proposta completamente differente. Nel 1896 A. Thiery propose che le distorsioni fossero correlate al modo in cui l'occhio e il cervello utilizzano la prospettiva per giudicare le distanze o le profondità. Thiery considerò le frecce di Müller-Lyer come disegni di un oggetto simile a un cavalletto visto in tre dimensioni; le gambe si allontanerebbero dall'osservatore nella figura ad angolo acuto mentre si avvicinerebbero in quella ad angolo ottuso. Eccettuata una breve discussione sulla « teoria della prospettiva » fatta da Robert S. Woodworth nel 1938, il suggerimento di Thiery è stato raramente preso in considerazione fino a tempi recenti.

Woodworth scrisse: « Nella figura di Müller-Lyer le parti oblique suggeriscono facilmente la prospettiva; se si segue la prospettiva, una delle linee verticali appare molto piú distante e perciò obiettivamente piú lunga dell'altra ». Questa citazione mette in evidenza le difficoltà immediate che si incontrano nello sviluppo di teorie adeguate lungo questa direttrice. La distorsione si verifica infatti anche quando si trascura il suggerimento della prospettiva, poiché le frecce appaiono generalmente piatte e tuttavia risultano distorte. Inoltre, nessun accenno viene fatto al meccanismo responsabile dei cambiamenti di dimensione. Una teoria adeguata basata sul suggerimento di Thiery deve mostrare perché la distorsione si verifica anche quando le figure appaiono piatte. Dovrebbe pure essere indicato il genere dei meccanismi cerebrali responsabili.

L'ipotesi che la prospettiva geometrica – la convergenza apparente delle linee parallele con la distanza – abbia un rapporto con il problema è convalidata dal verificarsi di tali distorsioni nelle fotografie di scene reali nelle quali la prospettiva è pronunciata. Due rettangoli di uguali dimensioni sono marcatamente disuguali se vengono sovrapposti a una fotografia di binari ferroviari convergenti (*si veda l'illustrazione a pagina 31*). Il rettangolo superiore nell'illustrazione, che sarebbe il piú distante se fosse un oggetto reale giacente tra

L'illusione di Ewald Hering. Le linee orizzontali sono naturalmente due rette parallele. I fisici e gli astronomi mostrarono un vivo interesse per le illusioni ottiche, preoccupati che le osservazioni visive potessero talvolta dimostrarsi inattendibili.

i binari, sembra piú grande di quello inferiore (apparentemente il piú vicino). Questa illusione è nota come « illusione di Ponzo ».

In maniera analoga, l'occhio tende a espandere l'angolo interno di una stanza, visto in una fotografia, e a restringere gli angoli esterni degli edifici (*si veda l'illustrazione a pagina 35*). L'effetto è esattamente lo stesso di quello delle figure di Müller-Lyer, che infatti assomigliano ad abbozzi di disegni di angoli visti in prospettiva. In entrambi i casi le regioni indicate dalla prospettiva come distanti vengono espanse, mentre quelle indicate come piú prossime risultano contratte. Le distorsioni avvengono in senso opposto alle normali contrazioni dell'immagine retinica che si hanno quando aumenta la distanza rispetto a un oggetto. Questo effetto è semplicemente fortuito, o è un indizio dell'origine delle illusioni?

Immagini paradossali

Prima di affrontare il problema di sviluppare un'adeguata teoria sulla prospettiva sarà utile considerare alcune curiose caratteristiche delle immagini normali. Le immagini sono il materiale tradizionale delle ricerche sulla percezione, ma tutte le immagini sono altamente artificiose e presentano problemi speciali al cervello che le percepisce. In un certo senso, tutte le immagini sono impossibili, essendo dotate di una realtà duplice. Sono infatti viste sia come configurazioni di linee giacenti su un fondo piatto sia come oggetti immersi in uno spazio tridimensionale del tutto diverso. Nessun oggetto reale può essere contemporaneamente sia bidimensionale sia tridimensionale, ma le immagini lo sono in qualche modo. Considerate come forme, sono viste come se fossero bidimensionali; considerate come rappresentazioni di oggetti, sono viste in uno spazio quasi tridimensionale. Le immagini perciò forniscono dati di ingresso visivi con caratteristiche paradossali. Sono per di piú ambigue, poiché la terza dimensione non è mai chiaramente definita.

Il cubo di Necker è un esempio di immagine in cui l'ambiguità della profondità è cosí grande che il cervello non riesce mai a selezionare un'unica risposta. Comunque, ogni proiezione in prospettiva potrebbe rappresentare una infinità di forme tridimensionali. Si potrebbe quindi pensare che il sistema percettivo abbia un compito impossibile! Fortunatamente per noi, il mondo degli oggetti non ha una varietà infinita; vi è sempre una possibilità massima e quindi riusciamo a interpretare le nostre immagini piatte piú o meno correttamente in termini del mondo degli oggetti.

La difficoltà del problema di scorgere la terza dimensione nelle due dimensioni di una immagine, o nelle immagini retiniche di oggetti normali, è ingegnosamente messa in evidenza da speciali « immagini impossibili » e « oggetti impossibili » che mostrano cosa accade quando all'occhio vengono presentate informazioni di distanza chiaramente incompatibili. Il triangolo impossibile inventato da Lionel S. Penrose e R. Penrose non può essere percettivamente interpretato come un oggetto nel normale spazio a tre dimensioni (*si veda l'illustrazione a pagina 36*). È tuttavia possibile costruire oggetti reali a tre dimensioni, e non semplici immagini, che provocano l'insorgere della medesima confusione percettiva quando sono visti con un solo occhio. Per esempio, il triangolo di Penrose può essere costruito come struttura tridimensionale aperta (*si veda l'illustrazione in alto a pag. 37*) che assomiglia a una struttura chiusa impossibile quando è vista con un solo occhio (o fotografata) da una determinata posizione (*si veda l'illustrazione in basso a pagina 37*).

Le immagini ordinarie non sono cosí differenti da immagini ovviamente impossibili. Tutte le figure che mostrano profondità sono paradossali: noi le vediamo sia come piatte (quali effettivamente sono) sia come aventi una sorta di profondità artificiale che non è del tutto esatta. Non ci viene la tentazione di toccare, attraverso la sua superficie, gli oggetti mostrati in una figura o quelli che sembrano sporgere? Cosa capita, però, se rimuoviamo la superficie? Il paradosso della profondità rimane?

La rimozione dello sfondo

Per rimuovere lo sfondo negli esperimenti di laboratorio si rendono lumi-

nose le immagini in modo che brillino nell'oscurità. Per privare il cervello dell'informazione stereoscopica, che potrebbe rivelare che le immagini sono effettivamente piatte, le immagini sono viste con un occhio solo. Possono essere composizioni di fili ricoperti con vernice luminescente o diapositive fotografiche illuminate posteriormente da un pannello elettroluminescente. In entrambi i casi non vi è sfondo visibile, in modo che si può valutare quanto lo sfondo è responsabile del paradosso di profondità delle immagini, comprese le immagini illusorie.

In queste condizioni le frecce di Müller-Lyer assomigliano a veri angoli in base alla loro prospettiva e possono perfino essere indistinguibili da angoli luminosi veri. Le figure non sono completamente stabili e talvolta capovolgono spontaneamente la loro profondità. Normalmente, però, appaiono come vuole la loro prospettiva e senza la profondità paradossale di immagini con sfondo. Le distorsioni sono ancora presenti. La figura che assomiglia a una freccia a due punte sembra un angolo esterno e appare contratta, mentre la figura con le punte della freccia rovesciate sembra un angolo interno e appare espansa. Ora, però, il paradosso è scomparso e le figure sembrano veri elementi d'angolo. Con un apparato adatto si potrebbe porre in rilievo la loro profondità come se fossero normali oggetti tridimensionali.

Avendo rimosso il paradosso, è possibile misurare, con mezzi abbastanza diretti, la distanza apparente di qualsiasi parte delle figure. Questo si può fare usando gli occhi come telemetro per misurare la profondità apparente della figura, che è visibile a un solo occhio. La figura illuminata dal retro è posta dietro un filtro polarizzatore mentre un secondo polarizzatore, posto ad angolo retto col primo, impedisce a uno degli occhi di vederla. Entrambi gli occhi però possono vedere una o più minuscole luci mobili di riferimento che sono introdotte otticamente nella figura mediante uno specchio semiargentato posto a 45 gradi con la linea di vista. La distanza di queste luci è data dalla visione stereoscopica, cioè dall'angolo di convergenza degli occhi; muovendo le luci in modo che sembrino coincidere con la distanza apparente delle parti prescelte della figura è possibile rappresentare su un diagramma lo spazio visivo dell'osservatore in tre dimensioni (*si veda l'illustrazione in alto a pagina 38*).

Quando questo diagramma è tracciato per vari angoli delle « ali » (cioè delle punte della freccia) della figura illusoria di Müller-Lyer, diventa chiaro che le figure sono percepite come angoli interni ed esterni. L'illusione della profondità si conforma strettamente ai risultati ottenuti quando si misurano indipendentemente le dimensioni dell'illusione chiedendo ai soggetti di scegliere, tra alcune linee di confronto, quelle che meglio si accordano con la lunghezza apparente della linea centrale dei due tipi di freccia (*si veda l'illustrazione in basso a pagina 38*). Nel secondo esperimento le figure vengono disegnate su uno sfondo strutturato normalmente, in modo che appaiano piatte.

I due esperimenti mostrano che quando lo sfondo è rimosso, la profondità segue molto strettamente l'illusione dei vari angoli. La somiglianza tra i risultati tracciati sul diagramma fornisce la prova di una importante e stretta connessione fra l'illusione che si ha quando la profondità non è vista e la profondità che viene vista quando è rimosso lo sfondo. Ciò significa che l'affermazione di Thiery era fondamentalmente esatta: la prospettiva in un modo o nell'altro può creare delle distorsioni. Ciò che risulta strano è che la prospettiva produce le distorsioni in accordo con la prospettiva *indicata* anche quando la profondità *non* è vista in maniera cosciente.

Costanza delle dimensioni

Il passo successivo è cercare qualche meccanismo percettivo che potrebbe produrre questo legame fra prospettiva e dimensioni apparenti. Un meccanismo percettivo adatto a questo ruolo, che sarebbe stato ovvio molti anni fa, è la costanza delle dimensioni. Questo fenomeno fu chiaramente descritto nel 1637 da René Descartes nella sua *Dioptrique*. Afferma infatti Descartes che « non è la misura assoluta delle immagini [negli occhi] che conta. Chiaramente esse sono 100 volte più grandi [in estensione] quando gli oggetti sono molto vicini di quanto lo siano quando essi si trovino 10 volte più lontano, ma esse non ci fanno vedere gli oggetti 100 volte più grandi. Al contrario, esse sembrano quasi della medesima misura, almeno finché non siamo ingannati da distanze troppo grandi. »

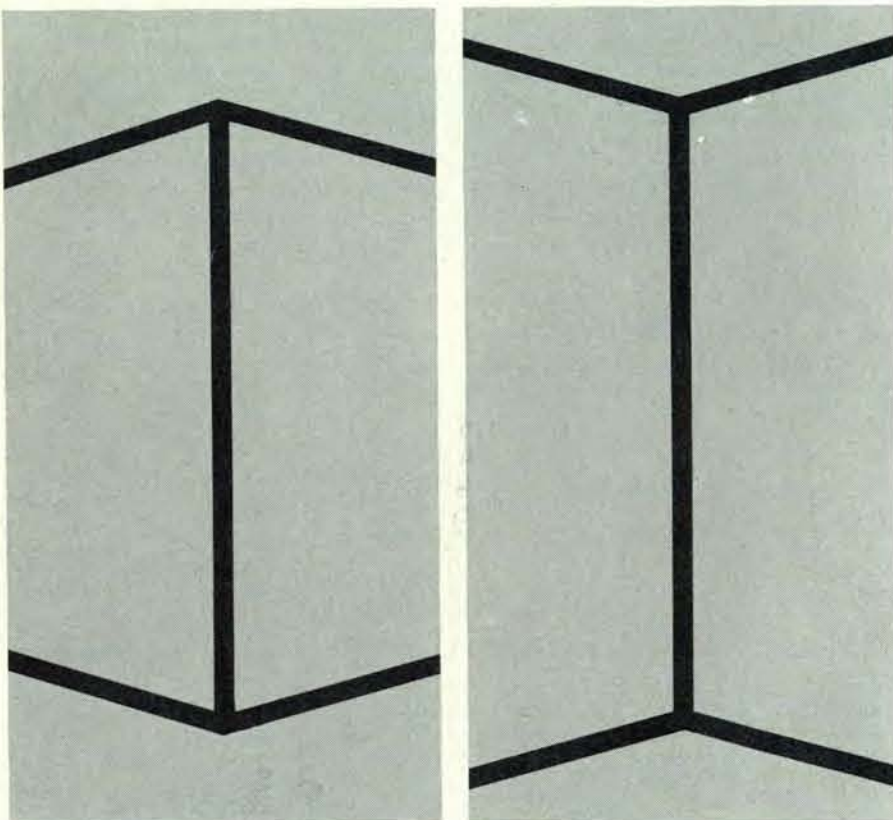
Sappiamo da molti esperimenti che Descartes aveva perfettamente ragione. Cosa si verifica, però, quando l'informazione di distanza, per esempio la prospettiva, è presentata all'occhio ma due componenti della scena, una delle quali dovrebbe essere ristretta dalla distanza, sono della stessa misura? Potrebbe darsi che la prospettiva presentata su una superficie piatta faccia agire il cervello in modo tale da compensare il previsto accorciamento delle immagini con la distanza anche se non vi è alcun accorciamento che occorre compensare? Se si verificano alcune di queste cose, è facile capire perché le figure che suggeriscono la prospettiva possono provocare l'insorgere delle distorsioni.

Ciò potrebbe costituire l'inizio di una teoria accettabile sulle illusioni. Le caratteristiche indicate come distanti dovrebbero risultare espanse, il che è proprio quanto constatiamo, almeno per le figure di Müller-Lyer e di Ponzo.

È probabile che questo modo di affrontare il problema non si sia sviluppato fino ai tempi recenti perché, sebbene la costanza dimensionale fosse da tempo perfettamente conosciuta, si è sempre supposto che essa seguisse semplicemente la distanza apparente in tutte le circostanze. Inoltre, non è stato sufficientemente capito in che modo figure molto strane costituiscano dati di ingresso visivi. Sono altamente atipiche e dovrebbero quindi essere studiate come casi speciali, essendo a un tempo paradossali e ambigue.

La costanza dimensionale è di solito identificata con un effetto conosciuto come legge di Emmert. Tale effetto può venire spiegato con un semplice esperimento implicante la dimensione apparente delle immagini residue che si hanno nella visione. Se si può ottenere una buona immagine residua (preferibilmente mediante una breve illuminazione della figura di prova con un flash elettronico), si può « proiettarla » su schermi o pareti posti a varie distanze. L'immagine residua può sembrare circa due volte più grande ogni volta che si raddoppia la distanza, anche se la dimensione dell'immagine prodotta dal flash rimane costante. È importante notare, in ogni caso, che vi è un cambiamento nella stimolazione retinica per ogni schermo o parete che giace a distanze diversa: le loro immagini *variano*. È possibile che il cambiamento delle dimensioni dell'immagine residua non sia completamente dovuto a un meccanismo secondo il quale il cervello cambia la propria scala in funzione delle sue dimensioni sulla retina rispetto alla dimensione dello schermo sul quale essa sembra giacere. Prima di procedere oltre, è essenziale scoprire se la legge di Emmert è dovuta semplicemente alla relazione fra le aree coperte dall'immagine residua e dallo schermo, o se l'informazione visiva della distanza cambia la dimensione dell'immagine residua per una qualche sorta di graduazione interna. Ci si trova tuttavia di fronte a un intricato problema sperimentale.

Come si viene a scoprire, vi è però una soluzione semplice. Possiamo usare il fenomeno di profondità ambigua del cubo di Necker per stabilire se la legge di Emmert è dovuta a una graduazione centrale da parte del cervello o è semplicemente un effetto delle aree relative di stimolazione della retina. Quando vediamo un cubo di Necker disegnato sulla carta rovesciarsi in profondità, non vi è un apprezzabile cambiamento di dimensione. Quando il cubo è mostrato su uno sfondo strutturato, esso occupa la paradossale profondità di tutte le immagini con sfondi visibili; non cambia dimensioni quando si rovescia nella pseudoprofondità.

Cosa accade, però, se rimuoviamo lo

L'illusione di Franz Müller-Lyer. Molte teorie furono successivamente invocate nel tentativo di spiegare come mai rovesciando le punte delle frecce (*a destra*) l'asta a esse connessa sembra allungarsi, mentre punte normali sembrano accorciarla (*a sinistra*).

La teoria dell'illusione di Müller Lyer, confermata dall'autore, suggerisce che l'occhio interpreti inconsciamente le figure a forma di freccia come strutture tridimensionali, simili a un angolo esterno (*a sinistra*) o a un angolo interno (*a destra*). Un meccanismo percettivo restringe la prima ed espande la seconda per compensare la distorsione provocata dalla prospettiva.

sfondo del cubo? L'effetto è sensazionale e completamente ripetibile: a ogni rovesciamento di profondità il cubo cambia la sua forma apparente, anche se non vi è alcun cambiamento nell'immagine retinica. La faccia che appare piú distante sembra sempre la piú grande. L'uso delle figure a profondità ambigua rende cosí possibile separare ciò che accade quando cambia il modello di stimolazione della retina. La risposta è che, almeno in parte, la costanza dimensionale, e la legge di Emmert, sono dovute a un meccanismo centrale di dimensionamento che risponde ai cambiamenti di distanza apparente anche se la stimolazione retinica rimane invariata.

La dimensione apparente viene quindi evidentemente stabilita in due modi. Può essere individuata con la sola distanza apparente. Può anche essere direttamente stabilita in base alle caratteristiche della profondità visiva, per esempio dalla prospettiva nelle figure a due dimensioni, anche se la profondità non è vista in quanto annullata dall'informazione di profondità che a essa contrasta, per esempio uno sfondo visibile. Quando sono presenti caratteristiche di profondità atipiche, la graduazione della dimensione è stabilita in modo errato e abbiamo una corrispondente illusione di distorsione.

La graduazione stabilita direttamente dalle caratteristiche di profondità (che dà luogo a distorsioni sistematiche quando è stabilita impropriamente) può essere definita « graduazione d'indizio-profondità ». È notevolmente costante e indipendente dall'« apparato » percettivo dell'osservatore. L'altro sistema è molto diverso e piú complesso essendo correlato solo indirettamente all'informazione retinica prevalente. È evidente che esso è collegato all'interpretazione dell'immagine retinica in termini dell'oggetto da essa rappresentato. Quando appare come un oggetto diverso, la graduazione cambia immediatamente per adattarsi all'oggetto alternativo. Se consideriamo la visione di un oggetto come un'ipotesi, suggerita (ma non strettamente provata) dall'immagine, possiamo chiamare il sistema « graduazione d'ipotesi-profondità » poiché esso cambia con ciascun cambiamento dell'ipotesi relativa a quale oggetto è rappresentato dall'immagine. Quando l'ipotesi è errata, abbiamo un'illusione che può essere straordinaria. Questi avvicendamenti nelle ipotesi stanno alla base dei cambiamenti in direzione, e anche in dimensioni, che si hanno quando si guarda l'ombra di una pala rotante.

Osservatori in movimento

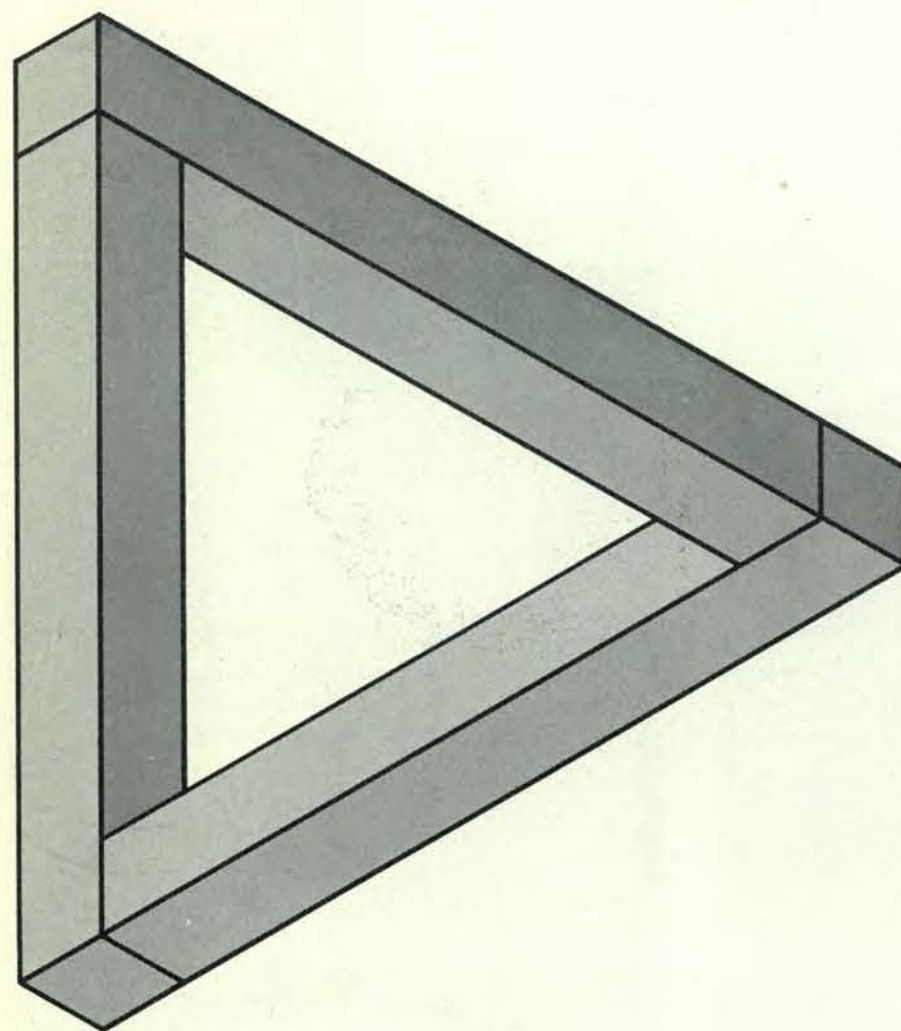
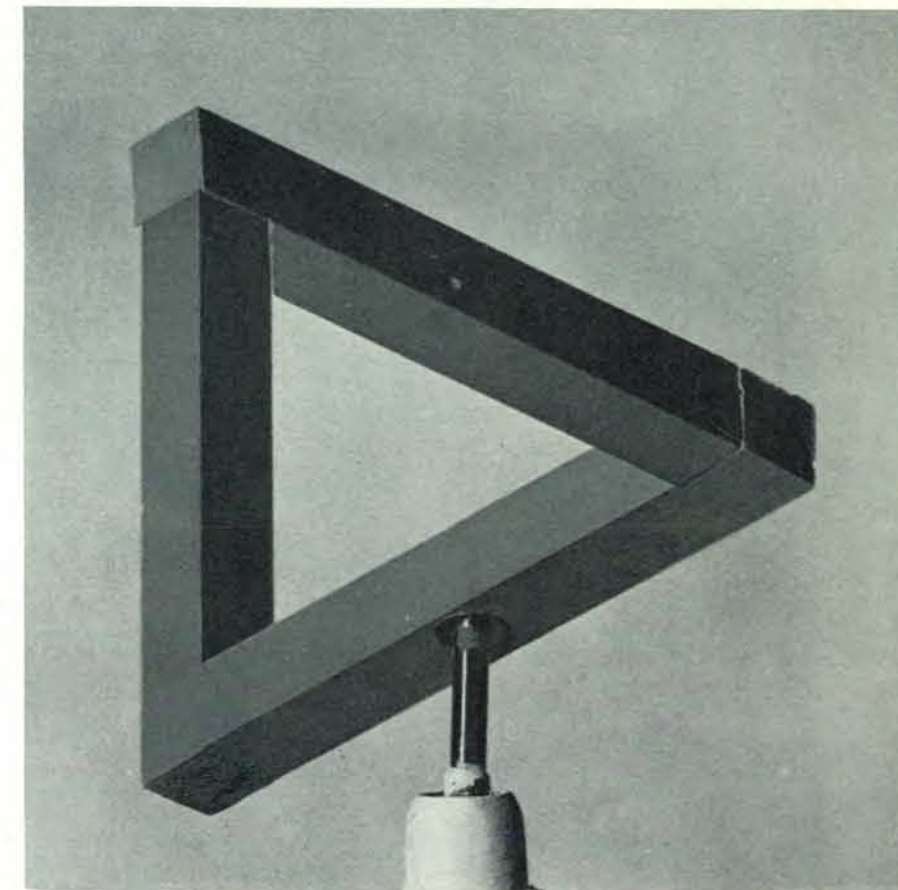
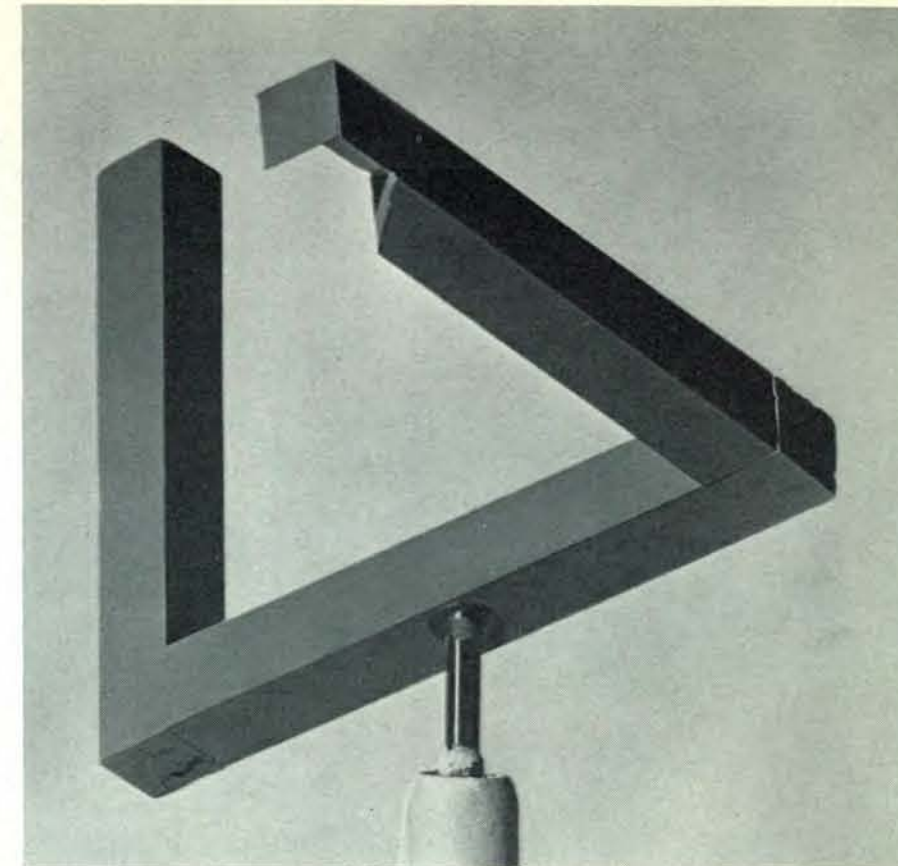
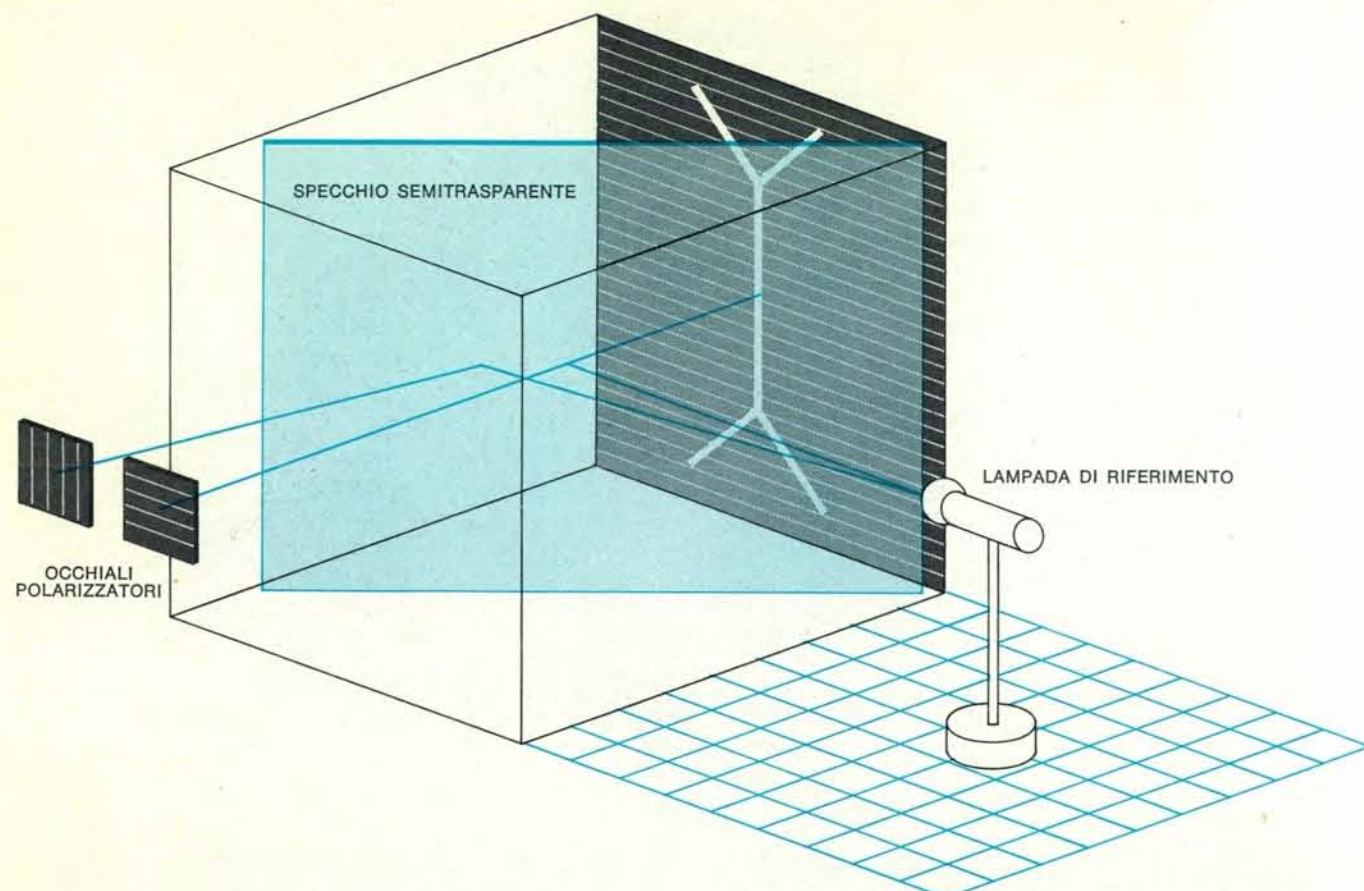
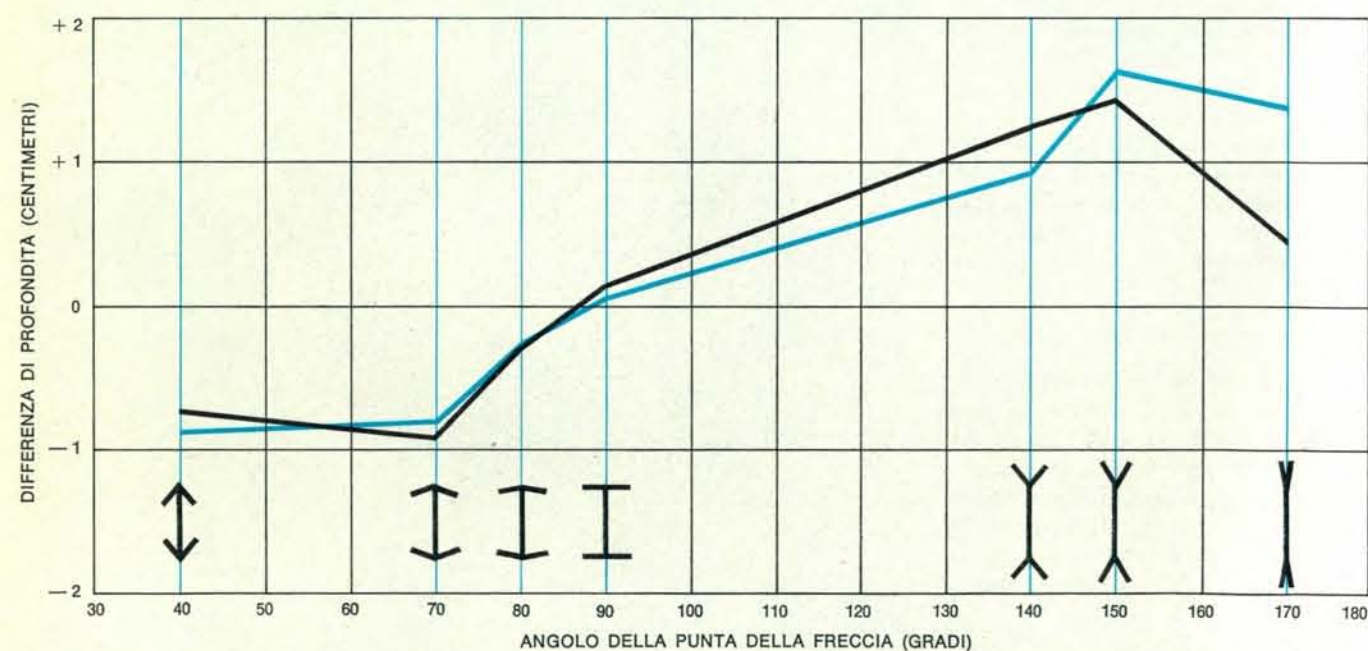
Le illusioni di distorsione tradizionali possono essere attribuite a errori nell'adattamento del sistema di graduazione indizio-profondità, che si originano quando le figure o gli oggetti hanno indizi di profondità ingannevoli, in particolare la prospettiva su una superficie piana. Per quanto queste illusioni possano talvolta infastidire i ricercatori che effettuano misurazioni visive, sono raramente un serio pericolo. L'altro tipo di illusione, l'inesatta graduazione dimensionale dovuta a un errore nelle ipotesi percettive predominanti, può essere serio in condizioni anormali o quando sono disponibili scarse informazioni visive, per esempio in un volo spaziale. Può anche essere importante quando si guida una macchina di notte o quando si deve far atterrare un aereo in condizioni di visibilità insufficiente. Le illusioni sono piú rischiose quando l'osservatore è in movimento rapido, perché allora anche un momentaneo errore può condurre a un disastro.

Finora non sono state eseguite molte misurazioni sulle illusioni sperimentate da osservatori in moto rispetto al loro ambiente. Le difficoltà sperimentali implicite nell'esecuzione di tali misurazioni sono notevoli; abbiamo però affrontato il problema in collaborazione con l'U.S. Air Force. L'apparecchiatura, che è abbastanza elaborata, può muovere l'osservatore con velocità e accelerazione controllate attraverso vari ambienti visivi, compresa l'oscurità dello spazio (con o senza stelle artificiali poste otticamente a distanza infinita).

La sensazione visiva di costanza dimensionale dell'osservatore durante il movimento viene misurata facendogli osservare un'immagine proiettata che cambia dimensioni quando egli si avvicina o si allontana da essa: al suo allontanarsi, è fatta espandere in grandezza, al suo avvicinarsi è fatta restringere. Il cambiamento delle dimensioni viene regolato finché l'immagine appare di dimensioni costanti all'osservatore in moto. Se non ci fosse alcun meccanismo percettivo per rendere costante la graduazione, le dimensioni di questa dovrebbero essere regolate in modo che l'immagine sulla retina dell'osservatore avesse sempre le stesse dimensioni, indipendentemente dalla distanza. Se, all'altro estremo, l'effetto di costanza dimensionale fosse completo, occorrerebbe lasciare invariata l'immagine che continuerebbe sempre ad avere le medesime dimensioni indipendentemente dalla sua distanza effettiva dall'osservatore. In pratica qualche cambiamento dimensionale fra questi due limiti fornisce l'illusione di una immagine invariabile e questo ci dà una misura dell'effetto della costanza delle dimensioni al muoversi dell'osservatore.

Abbiamo trovato che quando l'osservatore è in completa oscurità, si muove con velocità costante e osserva un'immagine proiettata dalle sue spalle su un grande schermo, non vi è un'apprezzabile costanza delle dimensioni. Quando è accelerato, la costanza delle dimensioni può manifestarsi ma può essere irregolarmente erronea. In particolare, se l'osservatore valuta il suo movimento in maniera errata, sia in direzione sia in entità, la costanza delle dimensioni viene di solito meno e può anche essere invertita. Ciò è abbastanza simile al rovesciamento della costanza delle dimensioni con il rovesciarsi della profondità nel cubo luminoso di Necker. Nei viaggi spaziali, la percezione può venire dominata dalle ipotesi prevalenti di distanza e velocità. Se sono entrambi errate, come può facilmente verificarsi quando si difetta di un'informazione visiva attendibile, l'astronauta può soffrire di illusioni ottiche che potrebbero risultare pericolose.

La parte non visibile della visione

La percezione visiva include la « lettura » nelle immagini retiniche di moltissime caratteristiche degli oggetti che non vengono rappresentate direttamente dalle immagini negli occhi. L'immagine non comunica direttamente molte caratteristiche importanti degli oggetti: se sono duri o soffici, pesanti o leggeri, caldi o freddi. Le caratteristiche non visive devono essere in un certo qual modo associate all'immagine visiva, per apprendimento individuale o probabilmente per via ereditaria, affinché gli oggetti siano riconosciuti mediante le relative immagini. Gli psicologi ritengono attualmente che l'apprendimento

Il triangolo impossibile fu inventato da L.S. Penrose e da R. Penrose dell'University College di Londra. Esso è logicamente coerente nei tratti parziali ma è globalmente senza senso. L'autore vede una certa analogia fra tali figure impossibili e le fotografie ordinarie, che danno l'illusione di una terza dimensione anche se sono piatte.

Il triangolo impossibile è stato in realtà « costruito » dall'autore e dai suoi collaboratori. L'unica limitazione è che deve essere visto con un solo occhio (o fotografato) ed esiste una sola posizione corretta per guardarlo. La fotografia in alto mostra come in realtà la struttura sia aperta, ossia due lati non si incontrino affatto; quella in basso, invece, scattata dalla posizione giusta, dà la perfetta illusione che essi si incontrino.

L'apparato per studiare le illusioni ottiche serve a presentare figure senza sfondo, affinché sembrino sospese. In tali condizioni le frecce di Müller-Lyer assomigliano a veri angoli. Il soggetto può variare la posizione di una piccola luce in modo che essa sembri giacere alla medesima profondità di una qualsiasi parte della figura. La luce, che il soggetto vede nello spazio tridimensionale con entrambi gli occhi, è sovrapposta alla figura illuminata tramite uno specchio semitrasparente. Un filtro polarizzatore è posto sopra la figura e il soggetto guarda attraverso gli occhiali polarizzatori che gli permettono di vedere la figura con un solo occhio. In tal modo egli non può dire se la figura è realmente bidimensionale oppure tridimensionale.

Dati quantitativi ottenuti dall'illusione delle frecce di Müller-Lyer. La curva in nero esprime i risultati medi di 20 soggetti ai quali fu chiesto di scegliere la linea di confronto che meglio si adattasse alla lunghezza dell'asta centrale a due punte disposte secondo gli angoli indicati. Quando le punte delle frecce formavano, con la linea verticale, angoli minori di 90 gradi, le linee di confronto erano piú corte di almeno un centimetro. Quando le punte erano disposte a 150°, la linea di confronto era piú lunga di 1,5 centimetri. La curva in colore mostra la massima differenza di profondità percepita per la stessa serie di frecce visualizzate, con lo sfondo rimosso, nell'apparato mostrato in alto. Le due curve si accordano abbastanza bene, salvo che nella disposizione estrema di 170°, in cui la figura non rappresenta piú veri angoli quando viene presentata nell'apparato.

percettivo individuale sia molto importante per l'associazione delle proprietà non ottiche degli oggetti alle relative immagini retiniche. Un apprendimento di questo tipo è essenziale per la percezione; in assenza di esso si avrebbe un semplice comportamento stimolo-risposta.

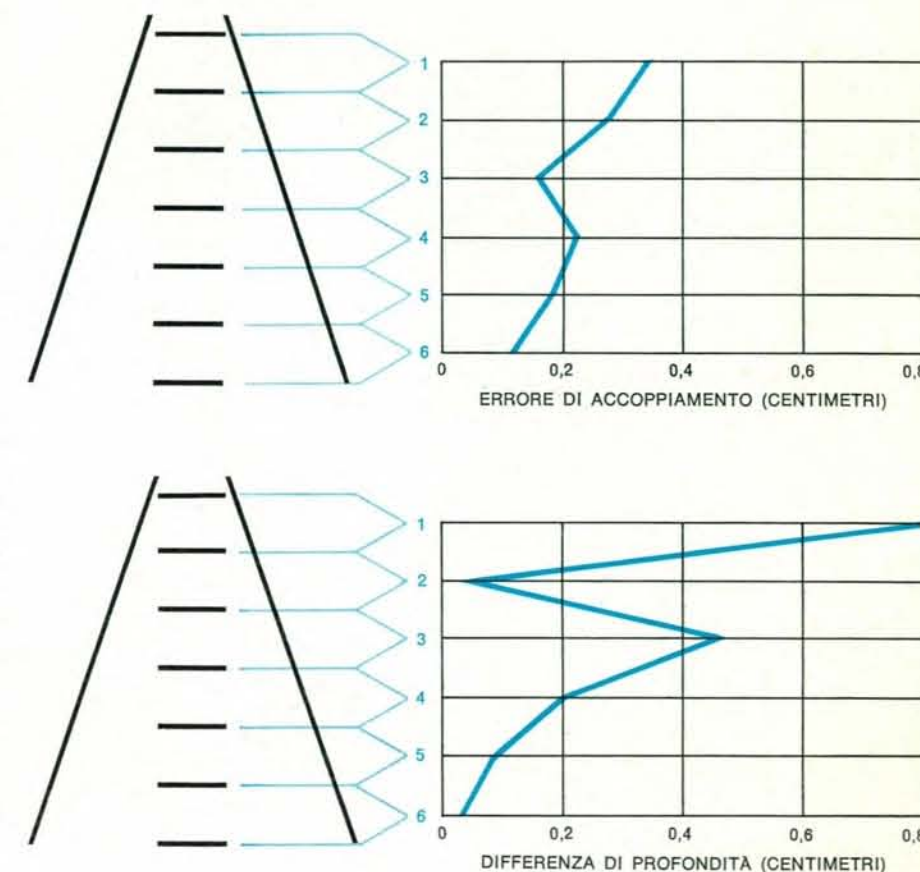
La percezione sembra essere un problema di consultazione delle informazioni già immagazzinate e inerenti agli oggetti e al modo in cui essi si comportano nelle varie situazioni. L'immagine retinica non fa altro che selezionare i dati memorizzati pertinenti. Tale selezione è abbastanza simile alla consultazione delle voci di un'enciclopedia: il comportamento viene determinato dalle voci anziché dallo stimolo che ha provocato la ricerca. Possiamo immaginare la percezione come essenzialmente costituita dalla selezione delle ipotesi memorizzate che meglio si accordano ai dati sensori in atto.

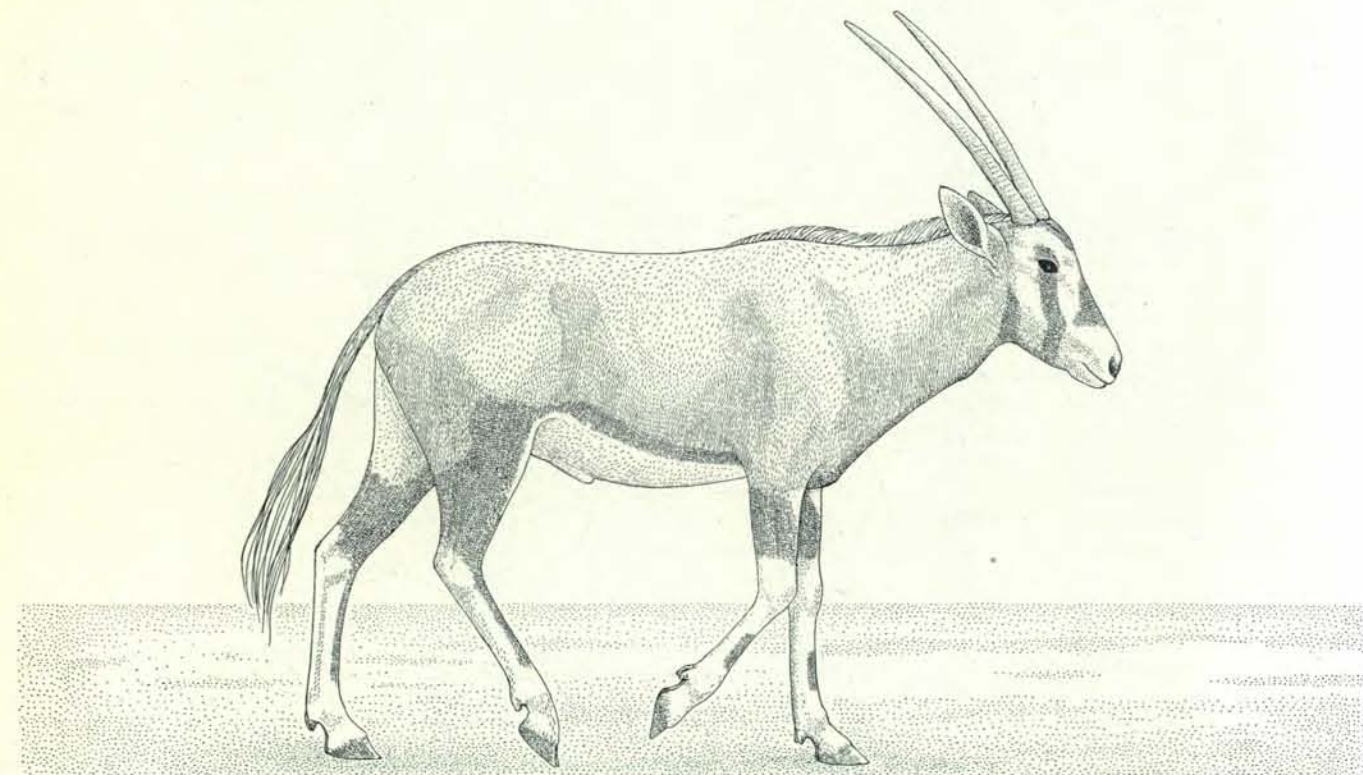
Un sistema di consultazione di questo tipo ha grandi vantaggi su un sistema di controllo che risponda soltanto ai dati di ingresso immediati. Se viene impiegata l'informazione immagazzinata, il funzionamento può continuare anche in temporanea assenza di informazioni pertinenti, o quando non vi sia un'informazione adeguata a fornire direttamente un preciso controllo del comportamento. Questo vantaggio ha importanti implicazioni per qualsiasi possibile sistema percettivo, compresa un'eventuale futura « macchina per leggere »: un robot attrezzato con occhi artificiali e un calcolatore e progettato per guidare veicoli o per manipolare oggetti mediante arti artificiali. Anche quando è disponibile una sufficiente informazione sensoria diretta per determinare le principali caratteristiche degli oggetti circostanti (il che avviene raramente), essa potrebbe richiedere una velocità di trasmissione dei dati superiore a quanto previsto dal sistema nervoso centrale (o da un calcolatore reale) per porre in grado un robot di comportarsi in maniera appropriata.

Vi sono perciò valide ragioni di carattere generale per supporre che un qualsiasi sistema effettivo di visione – o biologico o artificiale – dovrebbe usare normali informazioni sensorie per selezionare ipotesi preformate, o modelli, rappresentanti caratteristiche importanti degli oggetti del mondo esterno, e non limitarsi a controllare il proprio comportamento direttamente in base ai dati sensori di ingresso.

Se consideriamo i problemi connessi all'accumulo delle informazioni relative agli oggetti, appare subito chiaro che potrebbe risultare antieconomico immagazzinare un modello indipendente di ciascun oggetto per ciascuna distanza e orientazione che esso potrebbe occupare nello spazio circostante. Potrebbe essere di gran lunga piú economico memorizzare soltanto caratteristiche tipiche degli oggetti e usare le normali informazioni sensorie per adattare il modello selezionato alla situazione predominante. Il modello deve essere continuamente graduato per la distanza e l'orientazione se il possessore del sistema percettivo sta per interagire con l'oggetto reale.

Potremmo supporre che la graduazione d'indizio-profondità rappresenti tale adattamento del modello scelto alla luce delle informazioni di profondità disponibili. Quando l'informazione di profondità non è appropriata (come nel caso di caratteristiche prospettiche su una superficie piana), essa graduerà in maniera erronea il modello percettivo. Vi sarà un errore sistematico, ossia una illusione di distorsione dovuta a una graduazione non appropriata di profondità. Vi saranno pure errori – alcuni anche molto gravi – ogni volta che viene selezionato un modello erroneo. Possiamo renderci conto di ciò con una prova facilmente ripetibile riferendoci a certe figure ambigue, quali il cubo luminoso di Necker che cambia forma con ciascun rovesciamento di profondità, anche se le informazioni sensorie di ingresso permangono invariate.

Se questo modello dei meccanismi di percezione, basato essenzialmente su un sistema di consultazione, è corretto, potremmo aspettarci il sorgere di illusioni simili alle nostre in qualsiasi sistema percettivo reale, inclusi i futuri robot. Le illusioni non sono causate da qualche tipo di limitazione del nostro cervello. Esse sono il risultato delle imperfette soluzioni che è in grado di dare un qualsiasi sistema di manipolazione dei dati che affronti il problema di stabilire la realtà di oggetti partendo da immagini ambigue.

L'illusione dei binari ferroviari può essere studiata quantitativamente. I metodi sono identici a quelli descritti nell'illustrazione in basso nella pagina a fronte. Ai soggetti fu presentato un segmento orizzontale in una delle posizioni indicate e fu chiesto di sceglierne un secondo che sembrasse avere la stessa lunghezza. L'errore di accoppiamento è mostrato dalla curva in alto. Le coppie di segmenti furono quindi presentate nell'apparato mostrato in alto nella pagina a fronte, e i soggetti sistemarono la luce in modo da farla corrispondere alla profondità apparente di ciascuno. In tali condizioni (*curva in basso*), l'illusione di profondità è molto piú influenzata dalla posizione di una data coppia di segmenti, ma l'andamento della curva rimane inalterato.

L'antilope alcina è la piú grande antilope africana. Un maschio adulto pesa piú di mezza tonnellata e misura al garrese circa m 1,80. Appartiene alla famiglia dei Bovidi, è docile e può vivere anche in un habitat molto arido, inadatto al bestiame comune.

L'orice, altra grande antilope africana, ha al garrese un'altezza di m 1,20. È meglio acclimatata dell'antilope alcina nelle terre aride e si può trovare anche nelle zone desertiche; tuttavia l'orice è tutt'altro che docile e usa con prontezza le sue lunghe corna.

L'antilope alcina e l'orice

Queste grandi antilopi africane possono sopravvivere per un tempo illimitato senza bere in quanto sono dotate di meccanismi fisiologici che riducono al minimo la quantità d'acqua eliminata per evaporazione.

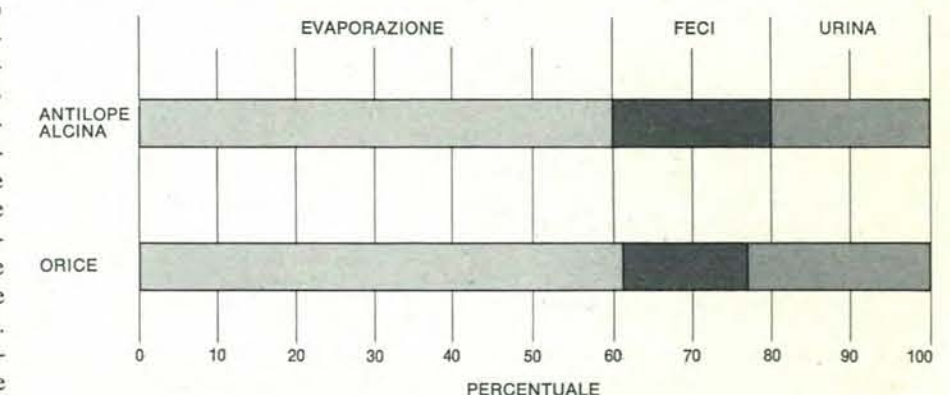
di C R. Taylor

Quando alcuni viaggiatori riferirono di aver visto la neve all'equatore e la notizia raggiunse per la prima volta la Londra ottocentesca, i dotti membri della Royal Geographical Society la posero in ridicolo. L'esistenza di antilopi nei deserti dell'Africa può sembrare a molti zoologi odierni tanto sorprendente quanto ai geografi dell'Ottocento la neve all'equatore. È infatti impensabile che in tale ambiente possa vivere un membro della famiglia dei Bovidi. Se poi si affermasse che questi animali sopravvivono senza bere, allora certi zoologi sostituirebbero « impensabile » con « impossibile ». Eppure è ora ben noto che la neve ricopre i picchi di molte montagne equatoriali, e naturalisti, cacciatori e indigeni affermano che certe antilopi del deserto possono sopravvivere per un tempo illimitato senza bere.

Per quale motivo la sopravvivenza nel deserto senza bere dovrebbe sembrare impossibile? È certamente impossibile per gli esseri umani; nel deserto, durante una giornata torrida, un uomo può perdere quasi 14 litri d'acqua a causa della sudorazione e dell'evaporazione, quindi non potrebbe sopravvivere senza bere. D'altra parte, il ratto canguro e altri roditori del deserto vivono senza bere, anche quando mangiano cibo secco; il loro bisogno di liquido è soddisfatto dalle minime quantità d'acqua che si trovano libere nei cibi di cui si nutrono e da quell'acqua che dalle sostanze alimentari si produce quando sono ossidate nei processi metabolici. Ma i roditori sono piccoli e possono sfuggire alle alte temperature del deserto scavando tane sotterranee. I grandi mammiferi non possono sfruttare questo tipo di riparo; per regolare la temperatura corporea durante la calura del giorno essi debbono evaporare notevoli quantità d'acqua. Anche il cammello, forse il piú conosciuto animale del deserto, ha il medesimo problema. Come fu scoperto dal fisiologo Knut Schmidt-Nielsen e dai suoi collaboratori, il cammello ha un'insolita capacità di limitare la sua perdita d'acqua per evaporazione; tuttavia, per poter sopravvivere, deve bere. Se anche le antilopi del deserto africano non avessero simili capacità, sarebbe impossibile la loro sopravvivenza.

Secondo quanto riferiscono numerosi testimoni oculari vi sono due antilopi – l'antilope alcina (o elano) e l'orice – che non hanno bisogno di bere acqua. L'antilope alcina (*Taurotragus*) è un grande, docile animale che occupa un certo numero di habitat dell'Africa orientale, ivi compreso il confine del Sahara. Si addomestica facilmente e è stata spesso proposta come l'animale che potrebbe utilizzare le terre che durante la siccità sono troppo aride per il bestiame comune.

Benché l'antilope alcina non abbia bisogno d'acqua, non si addentra nei deserti piú sterili; invece l'orice (*Oryx*) è veramente una specie del deserto; essa non cerca riparo neppure durante la calura del mezzodí, ma rimane esposta per tutto il giorno agli ardenti raggi solari. Come l'antilope alcina è docile, cosí l'orice è aggressiva; usa le sue corna a forma di spada con grande facilità e coloro che studiano la sua fisiologia debbono fare un esercizio fisico pari a quello mentale.

Alcuni anni fa, con l'aiuto di Charles P. Lyman dell'Università di Harvard, mi misi in viaggio per andare a vedere se queste due antilopi realmente vivano nei deserti africani senza bere e, se ciò è vero, come riescano a farlo. Mi ero posto tre semplici domande. Prima: l'antilope alcina e l'orice possiedono forse qualche insolito meccanismo per conservare l'acqua? Seconda: se cosí fanno, quanta acqua richiedono quando questo meccanismo è in funzione? Terza: possono rifornirsi dell'acqua che gli serve in un modo che non sia il bere? Ebbi modo di indagare su questi problemi alla East African Veterinary Research Organization, a Muguga nel Kenia, dove i direttori (in un primo tempo Howard R. Binns, poi Marcel Burdin) mi procurarono generosamente lo spazio necessario per il laboratorio e l'attrezzatura scientifica.

L'evaporazione attraverso la pelle e le vie respiratorie si è dimostrata il modo principale con cui le antilopi disperdono acqua quando in un ambiente desertico simulato hanno a disposizione tutta l'acqua che vogliono. L'unico modo di risparmiare quantità apprezzabili d'acqua nel calore di un deserto è quello di ridurre notevolmente l'evaporazione.

Il primo passo fu quello di preparare in laboratorio un ambiente che simulasse un deserto caldo e consentisse di calcolare tutta l'acqua eliminata dalle antilopi. Non tentai di riprodurre esattamente le stesse temperature minime e massime della notte e del giorno nel deserto, ma mantenni la temperatura media diurna per 12 ore e la temperatura media notturna per lo stesso periodo. Mediante un'apparecchiatura di riscaldamento e di condizionamento d'aria fui in grado di innalzare rapidamente la temperatura dello stabulario a 40 °C o di abbassarla a 22 °C, simulando cosí le temperature medie diurne e notturne del deserto.

In questo ambiente, misurai i vari modi in cui gli animali eliminavano acqua: trovai che la perdita per evaporazione era la piú importante (*si veda l'illustrazione a pag. 41*). La funzione esplicata dall'evaporazione è quella di evitare il surriscaldamento dell'animale. La maggior parte dei mammiferi mantiene la temperatura corporea a un livello pressoché costante, di solito attorno a 37 °C. Se la loro temperatura s'innalza troppo, essi muoiono; infatti per la maggior parte dei mammiferi una temperatura corporea di 43 °C che duri anche poche ore riesce quasi sempre mortale.

Quando la temperatura ambiente è inferiore a quella dell'animale, il che è quanto normalmente succede, il calore va dall'animale all'ambiente circostante (per conduzione e irraggiamento). Il mantenimento in queste circostanze di una temperatura corporea costante richiede che i meccanismi metabolici dell'animale producano una quantità di calore pari al flusso di calore verso l'esterno. Tuttavia, quando la temperatura ambientale è superiore a quella dell'animale, la direzione del flusso di calore si capovolge e esso passa dall'ambiente all'animale. Se in queste circostanze la temperatura corporea deve rimanere costante, il calore assunto dall'ambiente, come pure quello prodotto dal metabolismo, deve essere disperso per evaporazione. Ogni grammo d'acqua che evapora porta con sé 0,58 chilocalorie di calore: e in un uomo il sovraccarico di calore che si determina quando è esposto a un ambiente surriscaldato può essere tale da obbligarlo a eliminare per evaporazione piú di un litro d'acqua all'ora.

Un modo con cui il corpo può ridurre l'evaporazione sotto la sollecitazione del calore è quello di rinunciare a mantenere costante la temperatura corporea. Schmidt-Nielsen e i suoi collaboratori trovarono che quando il cammello ha a disposizione poca acqua, la sua temperatura corporea s'innalza durante il giorno di ben 7 °C. Per vedere se l'antilope alcina e l'orice avevano la stessa capacità, misurai nel laboratorio la loro temperatura rettale durante le 12 ore calde. Gli animali avevano a disposizione tutta l'acqua che volevano, sicché nulla impediva il mantenimento di una costante temperatura corporea mediante evaporazione. Ciononostante, durante le 12 ore a 40 °C la temperatura dell'antilope alcina s'innalzò di piú di sette gradi (da 33,9 a 41,2 °C) e la temperatura dell'orice di piú di sei gradi (da 35,7 a 42,1 °C) prima che l'aumento di evaporazione impedisse ogni ulteriore rialzo (benché di solito il rialzo della temperatura corporea fosse meno marcato). Cosí, invece di consumare acqua per mantenere costante la temperatura del corpo, gli animali « immagazzinavano » il calore nel loro corpo.

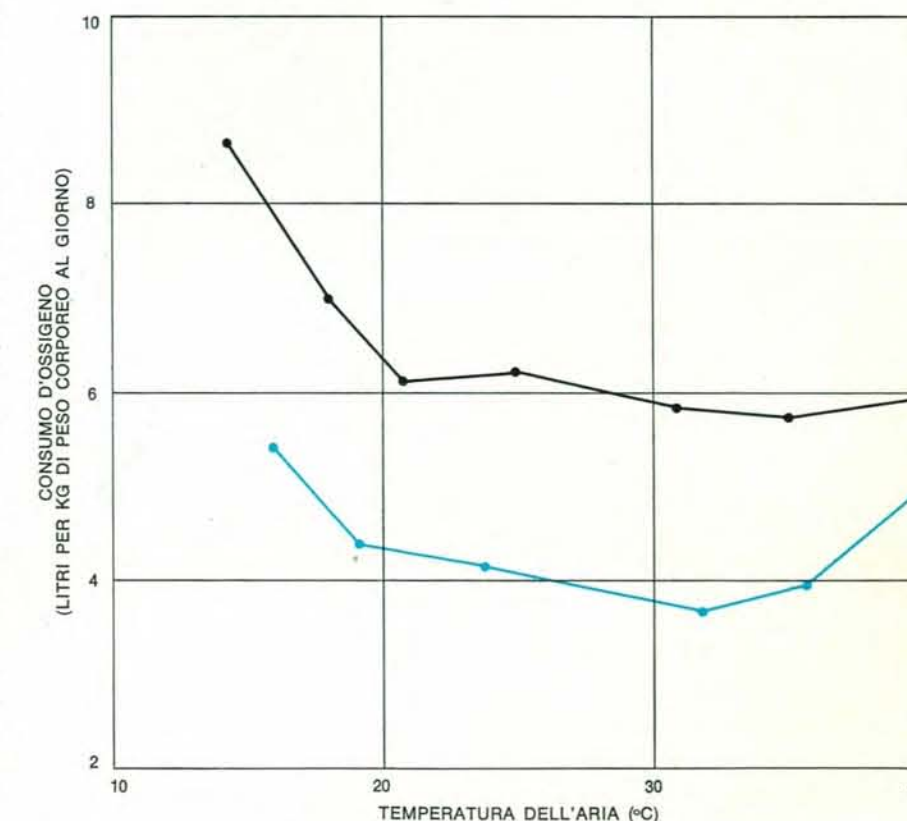
In un'antilope alcina del peso di 500 chilogrammi un aumento di temperatura di 7,3 gradi significa che l'animale è riuscito a immagazzinare circa 3000 chilocalorie. La dispersione dello stesso quantitativo di calore per evaporazione costerebbe piú di cinque litri d'acqua. Nel deserto, quando la temperatura notturna, piú bassa, permette un'inversione del flusso di calore dall'animale all'ambiente, questo calore immagazzinato viene disperso per conduzione e irraggiamento, piuttosto che per evaporazione.

Trovai che la temperatura corporea dell'antilope alcina e dell'orice, esposti a un'alta temperatura sperimentale e riforniti d'acqua, di solito s'innalzava di 3 o 4 gradi durante le 12 ore diurne. Dopo un'esposizione di 3 o 4 ore la traspirazione degli animali aumentava abbastanza per impedire un ulteriore rialzo della loro temperatura, che rimase al di sotto della temperatura del laboratorio anche alla fine delle 12 ore. Nel deserto questo modo di aumentare gradualmente la temperatura corporea prima che si metta in moto il meccanismo dell'evaporazione significa che gli animali possono passare le ore piú calde del giorno senza consumare preziosa acqua per abbassare la propria temperatura.

Queste prime osservazioni mostrarono che nelle antilopi adattate alla vita nel deserto, la perdita d'acqua per evaporazione veniva ridotta come nel cammello mediante l'aumento della temperatura corporea. Fin qui, tuttavia, avevo misurato le reazioni delle antilopi solo quando erano libere di bere quanta acqua volevano, ma ero ancora lontano dal verificare la loro capacità di sopravvivere senza bere del tutto. Perciò, prima di eseguire la successiva serie di misurazioni ridussi la fornitura d'acqua agli animali, finché essi non si disidratarono e persero peso. Venne data loro acqua sufficiente per mantenerli a un peso pari all'85 % del peso originario. Quindi, come nel precedente esperimento, li sottoposi alle stesse dodici ore calde diurne. La temperatura corporea dell'antilope alcina disidratata rimase al di sotto della temperatura esterna anche dopo 12 ore a 40 °C; naturalmente ciò può essere ottenuto solo mantenendo un elevato tasso di evaporazione. La temperatura dell'orice disidratata, invece, superava quella ambientale di un margine abbastanza largo da consentire la dispersione del calore metabolico per conduzione e irraggiamento; l'evaporazione non aumentava durante l'intera esposizione di 12 ore.

Sebbene avessi scelto 40 °C come temperatura dei periodi caldi, sapevo che la temperatura dell'aria del deserto è superiore a questo livello per qualche ora ogni giorno, e che i raggi solari a mezzogiorno nella zona equatoriale sono letteralmente infuocati. Durante la calura piú intensa del giorno l'antilope alcina si sposta all'ombra, ma l'orice sembra che non si curi dell'intenso calore. Le specie di orice col mantello di colore piú chiaro sono quelle che si spingono piú lontano nel deserto; il potere di riflessione del mantello deve certo ridurre il calore radiante. Tuttavia, a meno che il mantello non permetta una riflessione totale, un'orice che sta al sole del deserto deve assorbire molto piú calore di una che resta in laboratorio a 40 °C.

L'innalzamento di temperatura in un ambiente sperimentale produce la stessa risposta fisiologica da parte dell'antilope alcina (*a sinistra*) sia che venga abbondantemente rifornita d'acqua (*in nero*), sia che ne venga tenuta priva (*in colore*), come pure da parte dell'orice (*a destra*), quando viene abbondantemente rifornita d'acqua (*in nero*). A temperature dell'aria superiori ai 40 gradi, le temperature rettali degli animali non risultano piú superiori alle temperature ambientali (*la diagonale congiunge i punti con eguali temperature dell'aria e rettale*). Anzi, il raffreddamento dovuto all'evaporazione mantiene la temperatura rettale al di sotto della temperatura dell'aria, sicché il calore passa dall'ambiente all'animale. La temperatura dell'orice disidratata (*in colore*) continua ad aumentare, e il calore passa dall'animale all'ambiente; essa perciò non riceve calore dall'esterno.

Il raffreddamento per evaporazione dovuto piú al respiro ansimante che alla traspirazione è caratteristico dell'orice disidratata. Abbondantemente abbeverata (*in nero*), l'orice non solo suda ma incomincia ad ansimare appena la sua temperatura raggiunge i 39 °C; sopra i 41 °C l'orice disidratata (*in colore*) non suda, ma ansima vigorosamente.

Il consumo d'ossigeno dell'orice varia con la temperatura dell'ambiente e viene ridotto drasticamente dalla disidratazione (*in colore*). L'orice disidratata produce molto meno calore metabolico e la quantità di calore che va persa per evaporazione viene ridotta.

Decisi di saggiare le reazioni degli animali a un calore ancora piú insopportabile e innalzai la temperatura del laboratorio a 45 gradi. Disidratata o no, l'antilope alcina procurò di mantenere la sua temperatura circa cinque gradi al di sotto di questa nuova temperatura esterna e lo stesso fece l'orice quando fu rifornita d'acqua. Ma l'orice disidratata ebbe una reazione fisiologica completamente diversa. La sua temperatura s'innalzò fino a superare quella del laboratorio e rimase sopra i 45 gradi per ben otto ore senza evidenti segni di malessere. Credo che tale alta temperatura ininterrotta sia stata osservata soltanto in un altro mammifero: la piccola gazzella di Grant (*Gazella granti*). Questa ipertermia tanto nell'orice quanto nella gazzella le pone in grado di tenere in serbo grandi quantità d'acqua anche in condizioni di calore eccessivo e questo è probabilmente il fattore decisivo per la loro sopravvivenza nell'ambiente desertico.

Come fanno l'orice e la gazzella a sopravvivere a queste elevate temperature interne? Il cervello, con le sue complesse funzioni integrative, è probabilmente la parte del corpo piú sensibile alle alte temperature. È possibile che in entrambi gli animali il cervello rimanga notevolmente piú freddo del resto del corpo. L'arteria carotide esterna, che fornisce la maggior parte del sangue al cervello di questi animali, si divide in centinaia di piccole arterie parallele quando arriva nella regione chiamata seno cavernoso. Dalle vie nasali al seno defluisce sangue venoso meno caldo, riducendo presumibilmente la temperatura del sangue arterioso diretto verso il cervello. La prova di ciò è che, quando le temperature vengono registrate durante l'esercizio, il cervello di una gazzella risulta piú fresco di ben 2,9 °C rispetto al sangue arterioso che parte dal cuore. Mary A. Baker e James N. Hayward dell'Università di California a Los Angeles hanno dimostrato che anche la pecora è dotata di un simile meccanismo.

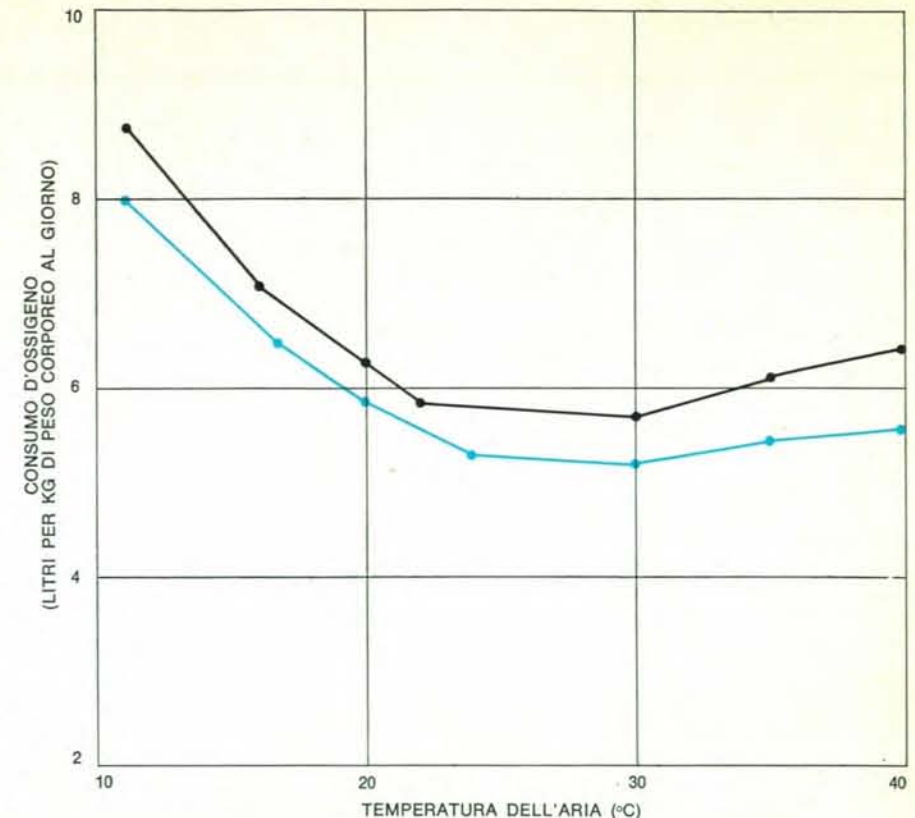
La maniera con cui un animale aumenta l'evaporazione per mantenersi fresco può fare diversa la quantità di calore che riceve da un ambiente caldo. Alcuni animali ansimano, altri sudano, altri ancora cospargono di saliva il corpo. Un animale che dipende dal sudore e dalla salivazione per eliminare il calore corporeo avrà necessariamente una temperatura cutanea inferiore a quella interna. Il sangue deve scorrere rapidamente verso la pelle, portando il calore interno alla superficie di evaporazione. Viceversa, un'alta temperatura cutanea e uno scarso afflusso di sangue riduce il passaggio di calore dall'ambiente caldo all'animale. Se l'animale può ansimare invece che sudare, riesce a sbarazzarsi del calore corporeo mediante l'evaporazione respiratoria e nello stesso tempo possiede una temperatura cutanea superiore che riduce al minimo l'accumulo di calore dall'ambiente. Volli perciò scoprire se l'orice, esposta a forte calore, aumentava l'evaporazione sudando, ansimando o con entrambi i sistemi. Mi domandai anche quale effetto potesse avere la disidratazione sull'importanza relativa delle due vie di evaporazione. Quando misurai il sudore e il respiro di un'orice liberamente rifornita d'acqua, trovai che la quantità di evaporazione aumentava ugualmente nelle due vie, ma che l'evaporazione cutanea era responsabile del 75 per cento del totale. Quando l'animale fu privato dell'acqua e esposto al calore, smise totalmente di sudare, ma cominciò ad ansimare quando la sua temperatura corporea superò i 41 °C.

Nel seno cavernoso, dove l'arteria carotide si ramifica in centinaia di vasi piú piccoli (*in colore*), avviene un raffreddamento controcorrente del sangue arterioso che dal cuore è diretto al cervello. Nel seno cavernoso il sangue venoso (*in nero*), proveniente dalle cavità nasali e raffreddato dall'evaporazione respiratoria, abbassa la temperatura del sangue arterioso. Il fatto che il cervello sia piú fresco del resto del corpo può essere di importanza capitale per la sopravvivenza nel deserto.

David Robertshaw dell'Hannah Dairy Research Institute in Scozia e io avevamo precedentemente scoperto che le ghiandole sudorifere dell'orice sono controllate da cellule nervose che liberano adrenalina; per esempio, si può far sudare un'orice per mezzo di una piccola iniezione endovenosa di adrenalina. Quando somministrammo le stesse dosi di adrenalina a un'orice disidratata, l'animale sudò ugualmente. Questo esperimento mette in evidenza che le ghiandole sudorifere di un'orice disidratata possono ancora funzionare, ma che il sistema nervoso dell'animale ha semplicemente cessato di stimolarle.

Gli aumenti di temperatura corporea e cutanea non sono gli unici stratagemmi che riducono al minimo l'evaporazione. Un abbassamento del metabolismo dell'animale potrebbe essere un altro modo con cui l'animale stesso riduce la sua produzione totale di calore. Per vedere se ciò era vero, misurai il metabolismo sia dell'antilope alcina sia dell'orice in una vasta gamma di temperature. Il metabolismo dell'alcina disidratata risultò leggermente ridotto, ma nell'orice la riduzione fu assai maggiore. A 40 gradi la riduzione che potei osservare nel metabolismo dell'orice permise un abbassamento dell'evaporazione del 17 per cento circa rispetto all'evaporazione da parte di animali liberamente riforniti d'acqua.

Presi insieme, questi dati indicano che, quando l'acqua è scarsa, l'antilope alcina e l'orice riducono in vari modi la loro evaporazione durante le caldissime ore diurne nel deserto. Entrambi gli animali immagazzinano il calore; entrambi riducono il passaggio di calore dall'ambiente, l'alcina cercando l'ombra e l'orice adattandosi alla massima temperatura corporea; entrambi riducono la quantità di calore metabolico prodotto.

Che cosa si può dire a proposito dell'acqua persa durante la notte? Quando il sole tramonta, le antilopi naturalmente non sono piú sottoposte alla sferza del calore ambientale. Il calore prodotto mediante i loro processi metabolici viene facilmente ceduto all'ambiente circostante, meno caldo, con modi diversi dal raffreddamento mediante evaporazione. Tuttavia una certa evaporazione continua di notte, sia attraverso le vie respiratorie sia attraverso la pelle.

La perdita d'acqua attraverso la pelle di notte non avviene col sudore: si tratta probabilmente di semplice diffusione. La pelle è leggermente permeabile all'acqua; anche i rettili, a pelle asciutta, perdono apprezzabili quantità d'acqua attraverso la pelle. Cosí è anche per l'antilope alcina e l'orice. Durante le 12 ore fresche del laboratorio, trovai che entrambi gli animali, quando avevano libero accesso all'acqua, perdevano attraverso la pelle circa mezzo litro d'acqua per ogni 100 chilogrammi di peso corporeo. La perdita d'acqua risultò ridotta quando gli animali vennero disidratati; la loro pelle sembrava piú asciutta e meno permeabile. La perdita d'acqua dell'alcina disidratata fu inferiore del 30 per cento rispetto a quando poteva bere liberamente. Nell'orice disidratata la perdita d'acqua attraverso la pelle risultò ridotta del 60 per cento circa.

Mi chiesi se anche la perdita d'acqua attraverso le vie respiratorie potesse essere ridotta. Quando un mammifero respira, l'aria inalata viene riscaldata nelle vie respiratorie fino alla temperatura corporea e saturata con vapore acqueo prima di raggiungere i polmoni. Normalmente la maggior parte dei mammiferi espira poi l'aria satura di umidità, ancora alla temperatura corporea. Donald C. Jackson della University of Pennsylvania School of Medicine e Schmidt-Nielsen hanno osservato, tuttavia, che due specie di piccoli roditori riescono a emettere aria molto piú fredda della temperatura corporea. È evidente che i roditori condensano una certa quantità di vapore acqueo lungo le vie respiratorie. Questo è un modo con cui le antilopi potrebbero ridurre al minimo l'evaporazione respiratoria. Due altri possibili modi di economizzare acqua sono in relazione al fabbisogno di ossigeno. Prima di tutto, se da ogni atto respiratorio può ricavare piú ossigeno, un animale non ha piú bisogno di far passare un gran volume d'aria attraverso le vie respiratorie, e cosí riduce la perdita d'acqua dovuta alla respirazione. Secondariamente, se il consumo d'ossigeno da parte dell'animale viene abbassato, anche il volume d'aria inspirata (e perciò la perdita d'acqua) viene ridotto.

Si sa che gli animali domestici emettono aria satura di vapore acqueo alla stessa temperatura del corpo. È proba-

Il consumo d'ossigeno dell'antilope alcina varia con la temperatura ambientale, ma viene poco ridotto dalla disidratazione (*in colore*). Come l'orice, l'antilope alcina aumenta il suo consumo d'ossigeno alla temperatura della notte: l'aumento minaccia entrambi di una perdita d'acqua se l'alimentazione notturna rende meno del 10 % d'acqua.

bile che l'antilope alcina e l'orice, con le loro grandi cavità nasali e con la respirazione relativamente lenta, facciano lo stesso, sicché essi non hanno a disposizione il meccanismo mediante il quale si può risparmiare acqua ricondensandola. Tuttavia entrambi gli animali hanno una temperatura corporea piú bassa di notte che di giorno e la differenza è sufficiente per ridurre in modo significativo la quantità d'acqua necessaria per saturare l'aria respiratoria. L'aria satura di umidità a 39 gradi, tipica temperatura diurna dell'antilope alcina, contiene circa 48 mg d'acqua per litro; l'aria alla tipica temperatura notturna di 33,8 gradi contiene circa il 25 per cento in meno di acqua.

Quando studiai gli animali alle temperature notturne, trovai che sia l'antilope alcina sia l'orice riuscivano a ricavare piú ossigeno dall'aria e respiravano piú lentamente quando avevano una temperatura corporea inferiore. Quando durante la notte l'antilope alcina ha una temperatura di 33,8 gradi, da ogni litro d'aria respirata essa riesce a ricavare il doppio di ossigeno. Come aumenta l'estrazione di ossigeno, aumenta anche la quantità d'aria inspirata a ogni atto respiratorio. Soltanto una parte dell'aria che l'animale respira raggiunge veramente il polmone, dove avviene lo scambio tra ossigeno e anidride carbonica; il rimanente riempie le vie respiratorie dove l'aria viene riscaldata e saturata di vapore acqueo, ma dove non avviene scambio di gas cioè nel cosiddetto «spazio morto». Quando l'alcina respira piú profondamente, il volume dello spazio morto rimane costante, ma una quantità di aria inspirata, proporzionalmente maggiore, raggiunge i polmoni: da ogni volume d'aria contenuto nei polmoni viene estratta la stessa quantità di ossigeno, ma il volume è stato aumentato. Ogni animale che respiri piú lentamente e profondamente estrarrà piú ossigeno dall'aria inspirata e perderà meno acqua (e calore) con l'aria espirata.

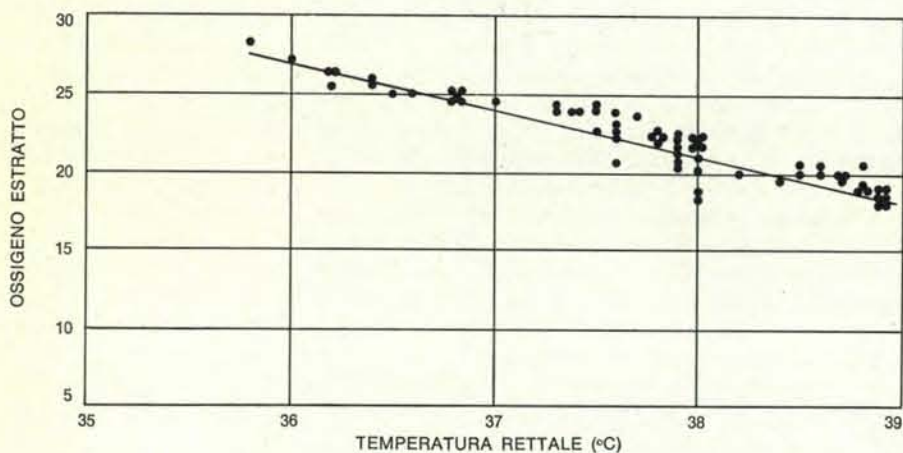
La velocità di respirazione di un'antilope alcina diminuisce quando la temperatura scende. Quando respira piú lentamente, essa respira anche piú profondamente, sicché una parte maggiore di aria raggiunge i polmoni, dove avviene lo scambio di gas.

Il ricavo di ossigeno da ogni respiro aumenta quando la temperatura dell'antilope alcina scende e l'animale respira piú lentamente e piú profondamente. Poiché il fabbisogno di O_2 è soddisfatto da un minor volume d'aria, la perdita d'acqua è ridotta.

Poteva avere importanza anche un minor consumo d'ossigeno? Avevo già scoperto che sia nell'antilope alcina sia nell'orice il metabolismo viene ridotto quando gli animali sono privati dell'acqua. Durante il periodo fresco, notturno, il metabolismo dell'antilope alcina disidratata era circa del 5 per cento inferiore rispetto a quello dell'animale che si poteva abbeverare, e il metabolismo dell'orice disidratata era inferiore di piú del 30 per cento rispetto a quello dell'animale che si poteva abbeverare. Mantenendo eguali le altre condizioni, quando nel deserto l'acqua scarseggia, la perdita d'acqua, per respirazione, delle due antilopi dovrebbe ridursi di una quantità proporzionale alla diminuzione del metabolismo. Benché sia inevitabile una certa perdita di acqua per respirazione, essa è ridotta al minimo dalla combinazione di tre fattori: abbassamento della temperatura corporea, aumento di estrazione dell'ossigeno e riduzione del metabolismo.

Quando misurai il consumo di ossigeno alle varie temperature, trovai che il metabolismo aumentava quando la temperatura scendeva sotto i 20 gradi. Ciò sembrava strano, poiché di notte le due antilopi incontrano spesso temperature inferiori ai 20 gradi. L'aumento, che serve soltanto a mantenere gli animali caldi, non sarebbe necessario se gli animali avessero una pelliccia leggermente piú spessa, che ridurrebbe anche l'aumento di calore durante le ore diurne. Sembra probabile che questo adattamento non si sia verificato perché l'aumento del metabolismo notturno segna la differenza tra una perdita netta o invece un risparmio d'acqua. Se un animale aumenta il metabolismo di notte, mangia anche di piú, introduce col cibo piú acqua libera e produce piú acqua per ossidazione. Nello stesso tempo, respirando piú aria per avere in maggior quantità l'ossigeno necessario, perde piú acqua. Quando si calcola la quantità d'acqua necessaria nel cibo per compensare la perdita netta d'acqua a causa dell'aumento del metabolismo, si ottiene che i calcoli tornano se nel cibo è contenuto un 10 per cento di acqua. Quindi, se di notte l'antilope alcina e l'orice mangiano piante con un contenuto di acqua superiore al 10 per cento, realizzano un guadagno d'acqua. Come vedremo, preferiscono piante che hanno un contenuto d'acqua notevolmente superiore a questa percentuale.

Trovato che i due animali possiedono meccanismi veramente insoliti per conservare l'acqua, cercai di stabilire quanta acqua – o piuttosto quanta poca acqua – fosse a ciascuno necessaria. Per far ciò tenni gli animali disi-

La sopravvivenza senza bere è possibile per l'orice e per l'antilope alcina poiché il loro cibo contiene quasi tutta l'acqua di cui hanno bisogno. Anche nei periodi di siccità le foglie di acacia, foraggio preferito dall'antilope alcina (*a sinistra*), contengono il 58 % d'acqua. Le foglie di un arbusto del genere *Disperma* e altri foraggi preferiti dall'orice (*a destra*) contengono poca acqua di giorno, ma possono contenere in media il 30 per cento d'acqua di notte. Cosí la quantità d'acqua che ciascun animale può ricavare nutrendosi (*in colore*) è piú di quella di cui ha bisogno per sopravvivere in un ambiente temperato (*in nero*) quando è disidratato e si avvicina alle quantità necessarie per la sopravvivenza in un ambiente desertico (*in grigio*).

dratati in due contrastanti ambienti di laboratorio: l'uno era il solito ambiente di 12 ore diurne calde e 12 notturne fresche; l'altro era costantemente fresco. Nell'ambiente fresco l'antilope alcina disidratata riuscí a mantenere un peso costante, pari all'85 % del peso originario, quando il suo attivo totale d'acqua (acqua libera nel cibo, acqua di ossidazione e acqua da bere) era appena al di sopra di 3,5 litri per 100 chilogrammi di peso corporeo per giorno. L'orice disidratata raggiunse appena la metà di questo quantitativo: un po' meno di 2 litri per 100 chilogrammi. In un regime di notti fresche e di giorni caldi, il fabbisogno d'acqua dell'antilope alcina crebbe a circa 5,5 litri e quello dell'orice a 3 litri. Questi dati mi indussero a pormi un'ultima domanda: potrebbero i due animali ricevere questa minima quantità d'acqua senza bere?

Le antilopi alcine non trovano soltanto ombra sotto gli alberi d'acacia; le loro foglie sono anche uno dei cibi preferiti. Raccolsi foglie d'acacia e misurai il loro contenuto d'umidità: anche dopo un periodo di notevole siccità esse contenevano in media il 58 % d'acqua. Calcolando il peso delle foglie di acacia che un'antilope alcina avrebbe mangiato in un giorno per far fronte al suo normale fabbisogno metabolico, trovai che le foglie avrebbero fornito circa 5,3 litri d'acqua per 100 chilogrammi di peso corporeo, ossia quasi esattamente la quantità necessaria all'antilope alcina disidratata per sopravvivere. Un'antilope che ricava questa quantità d'acqua soltanto brucando, probabilmente può vivere senza bere per un tempo illimitato.

L'orice preferisce erbe e arbusti, particolarmente un arbusto del genere *Disperma*. Durante le ore diurne queste piante sono cosí secche che le loro foglie cadono appena vengono toccate; secondo quanto potei misurare esse contengono appena l'uno per cento d'acqua. A prima vista questo sembrò una impossibile contraddizione. Tuttavia, pascolando c'è un modo per l'orice di procurarsi tutta l'acqua di cui ha bisogno.

Fin dal 1930 il naturalista inglese Patrick A. Buxton aveva osservato che di notte l'erba asciutta raccoglie umidità dall'aria del deserto, anche quando non c'è rugiada. La caduta della temperatura di notte, infatti, rialza l'umidità relativa e il tessuto vegetale secco può cosí assorbire umidità. Per determinare se questo meccanismo era importante per l'orice, esposi alcune piante all'aria condizionata del laboratorio, dove avevo creato la stessa temperatura e umidità media dell'aria notturna nel deserto. Nel giro di 10 ore le piante precedentemente inaridite avevano acquistato un contenuto d'acqua del 42 per cento.

Nelle zone selvagge, naturalmente, le piante non conterrebbero sempre questa stessa quantità d'acqua. Al tramonto il contenuto d'acqua sarebbe minore, ma piú tardi, di notte, le piante potrebbero essere sostanzialmente piú fredde dell'aria circostante a causa dell'irraggiamento verso il cielo notturno e cosí potrebbero raccogliere piú acqua delle piante usate per l'esperimento in laboratorio. È quasi certo che, mangiando principalmente di notte, l'orice può introdurre cibo contenente in media il 30 per cento d'acqua. Se questo accade, l'orice, che ha bisogno solo della metà dell'acqua rispetto all'antilope alcina, non sarebbe costretto a bere acqua mentre vaga per il deserto.

Sia l'antilope alcina sia l'orice hanno perciò adattamenti fisiologici e comportamentali fuori dal comune per poter vivere in un ambiente arido. Verrebbe quindi la tentazione di concludere che l'allevamento dell'antilope alcina e dell'orice in regioni aride potrebbe essere un modo eccellente per aumentare la produzione di carne in Africa. Un beneficio ulteriore sarebbe la conservazione di queste specie belle e interessanti. Tuttavia esistono seri problemi sul come immettere proteine d'antilope sul mercato a un prezzo di concorrenza rispetto a quello della carne di manzo (nella maggior parte dell'Africa Orientale la carne costa intorno alle 380 lire al chilo). Il buon esito dell'adattamento delle antilopi nelle regioni aride con vegetazione sparsa dipende dalla bassa densità di questi animali per chilometro quadrato; ciò rende difficile individuarle nelle vaste aree in cui vivono.

Per avere un buon prodotto dal punto di vista economico bisognerebbe sviluppare tecniche a buon mercato per catturare le antilopi, ucciderle, macellarle e trasportarle dai deserti alle città. L'alternativa all'allevamento allo stato brado – l'addomesticamento, la recinzione e l'alimentazione concentrata – vanifica i vantaggi fisiologici e comportamentali che le antilopi hanno sul bestiame comune. L'allevamento intensivo dei bovini da parte dell'uomo per la produzione di carne ha prodotto sotto questo aspetto un animale superiore alle antilopi. Solo gli economisti e gli agronomi possono risolvere questi problemi apparentemente insolubili dando inizio allo sfruttamento della carne di antilope.

La transdeterminazione nelle cellule

Una cellula embrionale che comincia a differenziarsi è determinata per una particolare funzione; tuttavia cellule larvali di Drosophila trapiantate su insetti adulti possono cambiare determinazione.

di Ernst Hadorn

Cercare di scoprire come il materiale genetico guidi lo sviluppo delle cellule e degli organismi, è sicuramente una delle piú appassionanti imprese della biologia moderna. Questa ricerca ha già fruttato molte importanti scoperte, ma è ancora ricca di fenomeni sorprendenti. Noi ci siamo imbattuti in uno di questi fenomeni durante esperimenti effettuati nel laboratorio dell'Università di Zurigo.

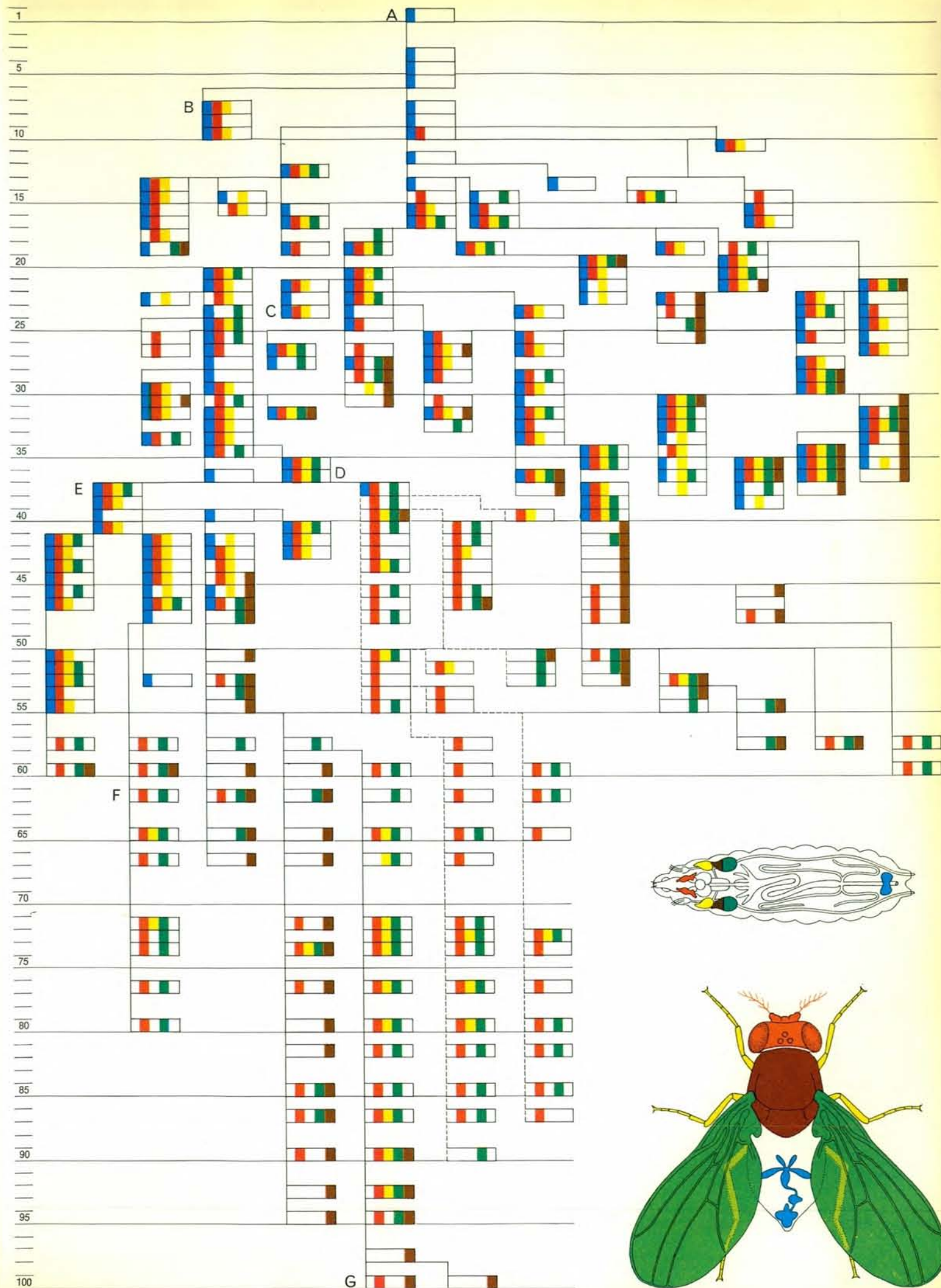
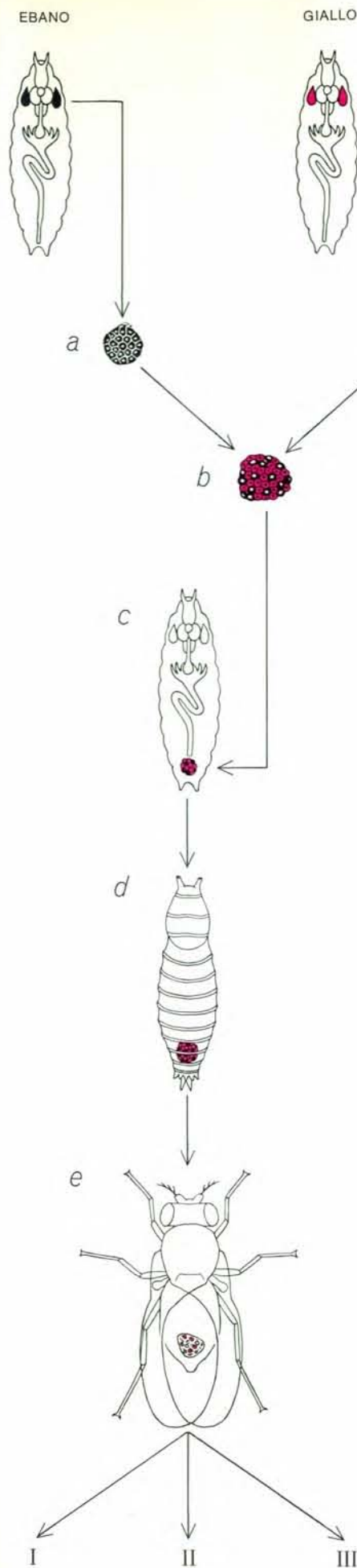
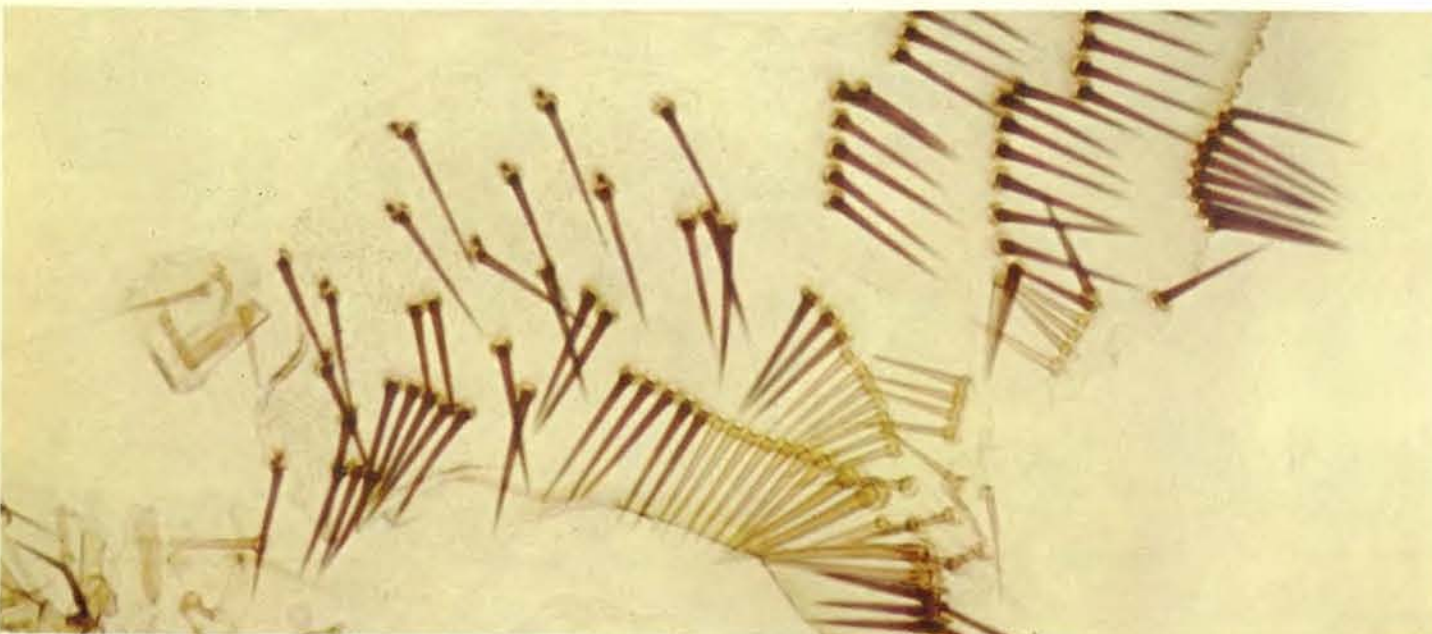
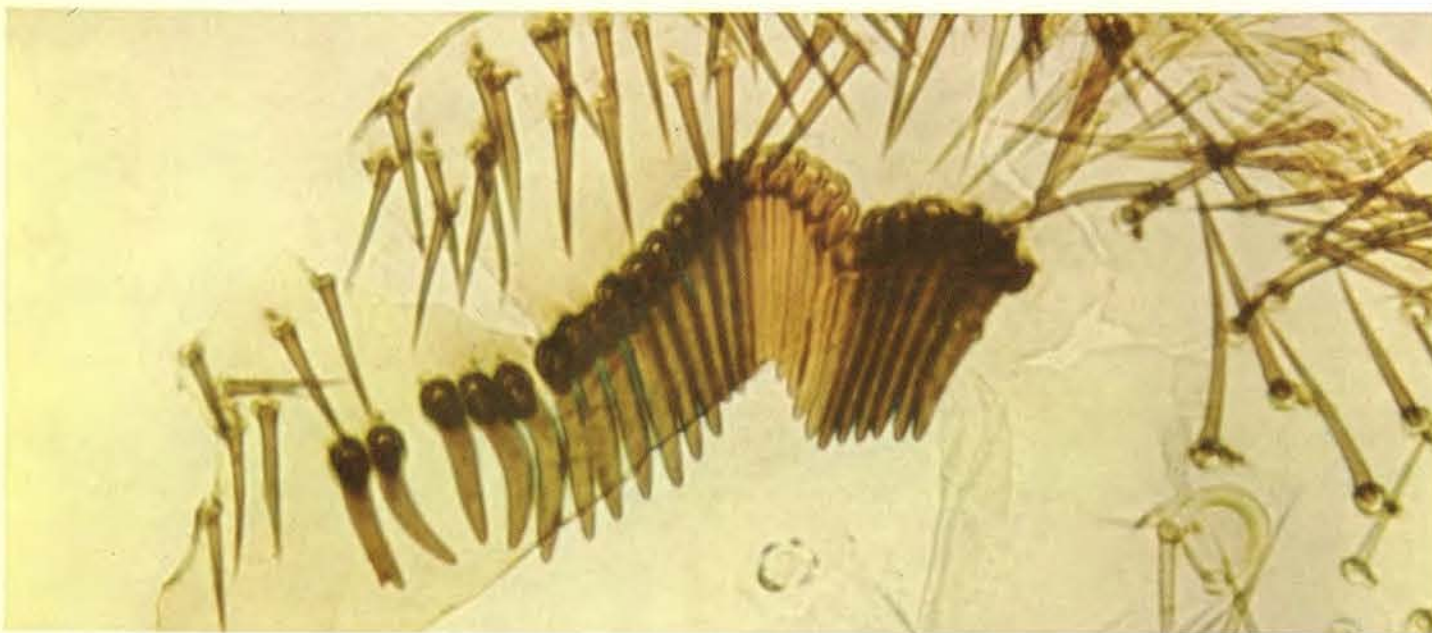
Vorrei prima di tutto spiegare quale fu l'idea da cui ebbero inizio i nostri esperimenti. Volevamo scoprire quando e in quali circostanze il futuro delle cellule embrionali viene determinato, cioè quando le cellule iniziano a specializzarsi e a formare le varie strutture differenziate dell'organismo adulto. Si sa che tutte le cellule embrionali partono con un identico corredo di geni che contengono informazioni per lo specifico sviluppo e funzionamento di ogni genere di cellula che si trova nell'organismo. Si ammette generalmente che lo specifico modo di differenziarsi di una cellula venga determinato quando sono attivati particolari gruppi di geni dei quali alcuni dirigono la differenziazione in cellule muscolari, altri in cellule nervose, altri in cellule dell'intestino e cosí via: tali cellule si vengono a trovare in un certo stato « determinato ». Perciò la determinazione è un evento importante, per il quale cellule di identico genotipo incominciano a essere diverse nelle loro caratteristiche, cioè nel fenotipo. In quale momento dello sviluppo embrionale si instaura questa determinazione? E quanto rigidamente è fissato tale stato di determinazione? Potrà essere ereditato da tutte le cellule che derivano da una cellula progenitrice già determinata o le cellule potranno talvolta essere deviate verso nuove strade?

L'organismo che abbiamo scelto per i nostri esperimenti è il moscerino della frutta *Drosophila melanogaster*. Questo insetto, che ha reso grandi servigi ai genetisti, sembrerebbe a prima vista assai meno adatto degli embrioni di anfibi o di ricci di mare per gli studi sullo sviluppo. Per i nostri scopi, tuttavia, la *Drosophila* si prestava assai bene.

Negli insetti piú evoluti, come mosche, api e scarafaggi, avviene una netta separazione delle cellule già ai primi stadi dello sviluppo embrionale. In una categoria di cellule embrionali la specializzazione finale procede senza indugio, cioè viene iniziata non appena le cellule incominciano a formarsi a seguito del processo di segmentazione e di formazione dei primi abbozzi di organi. Queste cellule formeranno il corpo della larva dell'insetto, con tutti i suoi organi. Un'altra categoria di cellule viene isolata a parte in quelli che sono chiamati dischi o « bottoni » imaginali. Sebbene le cellule dei dischi imaginali siano a contatto con le cellule vicine già differenziate, rimangono nello stato embrionale durante tutto il periodo larvale. In questo primo periodo della vita dell'insetto le cellule dei dischi cominciano a dividersi. Perciò quando la larva è pronta a trasformarsi in pupa e a iniziare il processo di metamorfosi, in ogni disco sono presenti alcune migliaia di cellule. Nella figura della pagina accanto si può vedere la collocazione di alcuni dischi imaginali nella larva di *Drosophila*.

Durante la metamorfosi la maggior parte degli organi larvali regredisce dentro l'involucro della pupa e alla fine le loro cellule si disintegrano. Nello stesso tempo, e solo in questo momento, le cellule dei dischi perdono il loro carattere embrionale e si differenziano nei diversi tessuti del corpo dell'imago o insetto adulto. Ciascuno dei dischi contribuisce alla formazione dell'insetto completo, come si può vedere in basso a destra nella figura della pagina a fronte. Un disco, quello genitale, contiene cellule che daranno origine alle strutture degli organi sessuali, alle parti posteriori dell'addome e all'intestino posteriore. C'è un disco per ciascuna delle sei future zampe e tre paia di dischi si uniscono a formare il capo (di questi ultimi dischi, solo quello occhio-antenna compare nella figura).

Ciascun disco forma un insieme di diverse strutture. Il disco della zampa, per esempio, contiene cellule che si svilupperanno rispettivamente in uncini, parti del tarso, tibia, femore, trocantere,

Cellule prelevate da una larva del moscerino della frutta, *Drosophila*, sono state mantenute vive e indifferenziate trapiantandole nell'addome di moscerini adulti ospiti, per 100 generazioni successive; le alterazioni del controllo genetico delle cellule sono rappresentate nel diagramma della figura nella pagina accanto. Queste cellule sono state tolte da un ammasso di cellule, o disco imaginale, destinato a formare, nel moscerino adulto, gli organi genitali, parti dell'intestino e dell'addome (*area blu nella larva e nel moscerino*). A ogni trasferimento alcune delle cellule furono trapiantate in larve che si metamorfosarono insieme alle cellule trapiantate. All'ottavo trasferimento (*B*) i controlli genetici delle cellule cominciarono a cambiare: la metamorfosi produsse parti della testa (*rosso*), parti delle zampe (*giallo*) e, non molto tempo dopo, parti delle ali (*verde*) e del torace (*marrone*). Dalla 56ª generazione in poi nessuna delle cellule di qualsiasi discendenza diede piú origine a organi genitali. Le parti del diagramma contrassegnate con le lettere da *A* fino a *G* si riferiscono alle strutture delle micrografie a pag. 55. Le discendenze unite da linee tratteggiate presentano anormalità dovute a mutazioni.

coxa e le parti di congiunzione col torace. Inoltre in ciascuna sezione della zampa debbono essere svolte molte funzioni diverse; cosí setole e peli di varie taglie e forma si sviluppano da differenti cellule distribuendosi secondo un preciso disegno.

I dischi imaginali possono essere isolati da una larva d'insetto. Nei nostri esperimenti sulla *Drosophila* li abbiamo isolati con un sottile ago di tungsteno e li abbiamo trapiantati nella cavità del corpo di altre larve. Quando le larve adulte si trasformarono in pupe, anche i trapianti si differenziarono in strutture adulte. Per esempio, quando un pezzo del disco trapiantato conteneva cellule determinate per gli occhi, nell'addome dell'ospite si sviluppava un occhio completo. Dopo molti esperimenti abbiamo trovato che ciascun disco imaginale contiene un mosaico di diverse popolazioni di cellule; questi dischi sono determinati per dar luogo a particolari strutture entro particolari regioni del corpo. Alcune regioni del mosaico del disco genitale maschile, per esempio, si svilupperanno nel dotto eiaculatore, altre regioni si differenzieranno nei vari elementi del pene, altre ancora daranno origine alle placche anali e all'intestino posteriore. In breve la futura differenziazione delle cellule è già determinata allo stadio larvale.

Ma come si comporta la singola cellula? Può essere data la prova che ciascuna cellula di una regione a mosaico di un disco imaginale è predestinata per una particolare differenziazione, indipendentemente dalle cellule con le quali si troverà a contatto piú tardi?

Abbiamo studiato questo problema riducendo i dischi in singole cellule e poi trapiantando una miscela di cellule provenienti da dischi diversi in una larva ospite (*si veda il disegno a sinistra*). Il moscerino della frutta è particolarmente adatto per questo tipo di esperimenti poiché vi è un'ampia disponibilità di mutanti facilmente riconoscibili, che cioè differiscono l'uno dall'altro per caratteristiche visibili come il colore del corpo e la forma delle setole. Perciò potemmo mescolare cellule provenienti, per esempio, da dischi di un mutante giallo e di uno nero ebano, lasciare che le cellule si aggregassero e poi vedere dove ciascuna cellula era andata a finire quando il trapianto aveva formato le strutture adulte nella larva ospite.

Risultò che le singole cellule dei dischi sono davvero già programmate per una differenziazione specifica. Esse collaborano a formare strutture normali. Per esempio, le cellule gialle e quelle color ebano si combinano a formare segmenti di zampe che sono un mosaico di cellule gialle ed ebano (*si vedano le microfotografie nella pagina a fronte*). I miei collaboratori Rolf Nöthiger, Heinz Tobler e Antonio Garcia-Bellido portarono a termine molti esperimenti con miscele di cellule di dischi imaginali in varie combinazioni. I risultati dimostrarono che in qualche modo, durante lo sviluppo del moscerino ospite, quelle cellule della miscela trapiantata che portano un particolare programma di differenziazione si riconoscono l'una con l'altra e si muovono fino a riunirsi a formare precisamente quella struttura. Al contrario, le cellule che provengono da dischi di regioni diverse non si mescolano mai, ma rimangono separate le une dalle altre. Una cellula che proviene da un disco della zampa non si mette mai insieme a un'altra che proviene da un disco delle ali, e le cellule della base delle ali non collaborano con le cellule dell'apice a formare la base o l'apice.

Potemmo perciò concludere che ciascuna cellula di un disco imaginale di una larva del moscerino della frutta è individualmente determinata per una differenziazione specifica. Questa determinazione si instaura molto prima che si possa rivelare una differenza morfologica delle cellule. Negli esperimenti che ora verranno descritti, abbiamo lavorato con cellule di dischi genitali già determinate, ma non ancora differenziate.

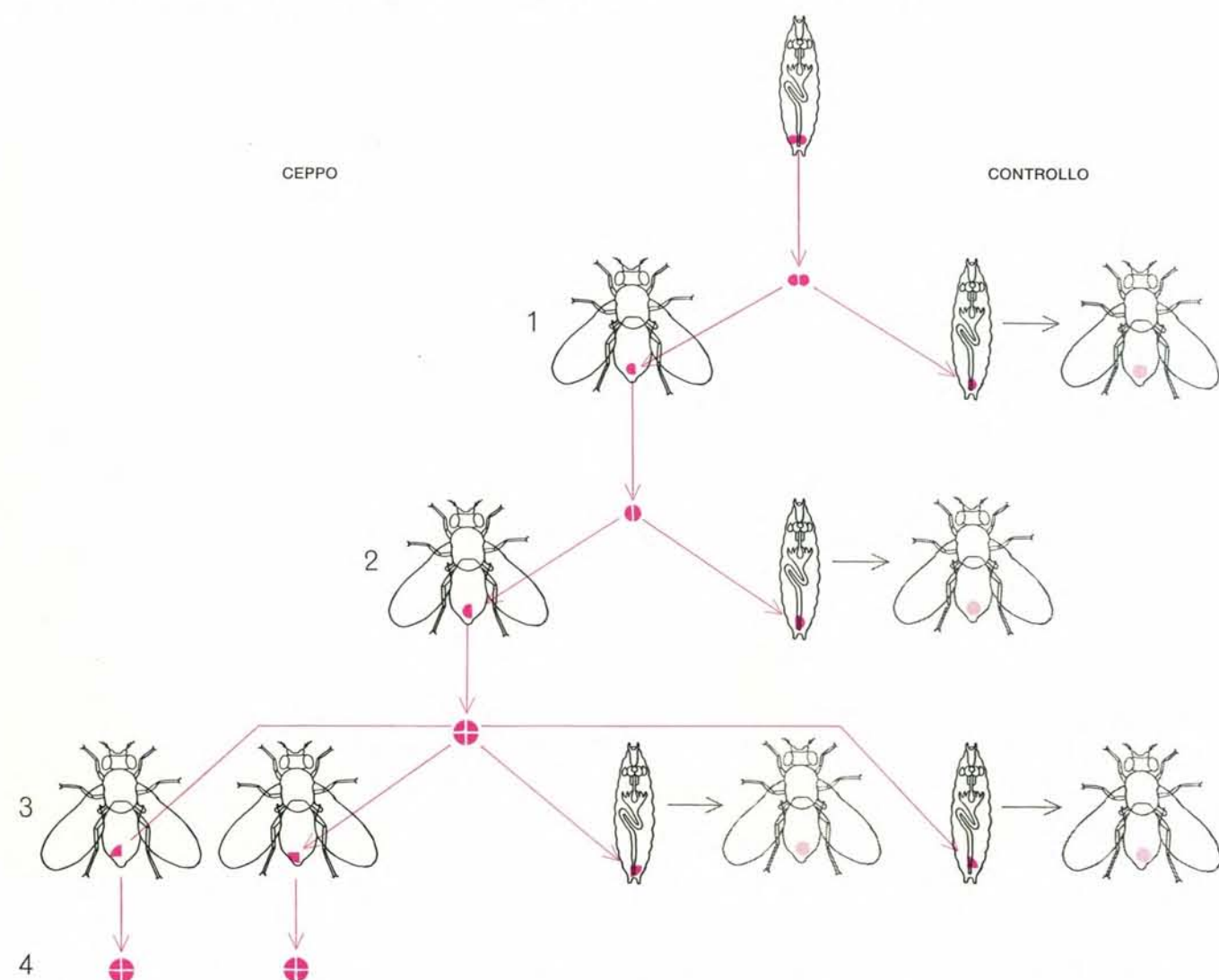
Quando incominciammo questa ricerca, ci siamo domandati cosa sarebbe accaduto se avessimo trapiantato parti di dischi larvali direttamente su moscerini adulti. Con tale procedura si

Con parti estratte da dischi imaginali delle zampe di larve del moscerino della frutta, appartenenti a razze di colore diverso (*a, in alto nel disegno*) viene fatta una miscela di cellule della zampa. Le porzioni dei dischi sono poi ridotte a cellule singole, mescolate insieme (*b*) e trapiantate nella cavità del corpo di una larva ospite (*c*). La larva ospite si trasforma in pupa (*d*) e le cellule trapiantate maturano. Quando vengono rimossi dal moscerino adulto (*e*) gli organi della zampa sono in parte gialli e in parte color ebano (*si vedano le microfotografie a fronte*). Tra questi organi si osservano (*I, fotografia in alto*) una parte terminale di zampa con un uncino ebano e uno giallo, (*II, al centro*) una cresta genitale ebano e gialla e (*III, in basso*) setole tibiali gialle e ebano. La mescolanza dei colori dimostra che le cellule con la stessa differenziazione potenziale si trovano le une con le altre e si uniscono a formare organi normali.

sarebbe scavalcato lo stadio ninfale. Quando tentammo l'esperimento ci trovammo di fronte a tre risultati del tutto inaspettati.

Primo: quando le cellule larvali vennero trapiantate nell'addome di un moscerino adulto, esse si divisero e si accrebbero senza alcuna limitazione. In condizioni normali di sviluppo le stesse cellule avrebbero cessato di dividersi non appena si fosse iniziata la metamorfosi pupale. Sotto l'influenza dell'ormone ecdisone, o « ormone della muta », esse avrebbero incominciato a differenziarsi nelle strutture adulte come setole, peli, uncini e cellule ghiandolari. L'accrescimento illimitato di queste colture continua ormai da sei anni. Siccome il moscerino vive per circa un mese, abbiamo dovuto trasferire le colture in moltiplicazione ogni due settimane circa. Le colture piú vecchie sono ora al centosessantesimo trapianto. Per il mantenimento permanente di una coltura dovevamo solo prendere una piccola parte di un trapianto per dare inizio alla successiva generazione. È cosí possibile avere molte subcolture da una singola coltura.

Il secondo risultato fu ancora piú sorprendente. Sebbene le cellule delle colture abbiano ormai vissuto per anni in moscerini adulti, esse conservano i loro originali caratteri embrionali. Non fu mai osservata nessuna differenziazione; evidentemente la fase di pupa è indispensabile per l'inizio della differenziazione. Da quanto abbiamo osservato possiamo concludere che l'ecdisone, che ha un ruolo dominante durante la metamorfosi pupale, non è piú presente, o non è attivo, nell'imago, cioè nell'insetto adulto. I liquidi del corpo dell'insetto adulto, nel quale le cellule della coltura vivono, sembrano essere del tutto simili ai liquidi del corpo dell'insetto allo stato di larva.

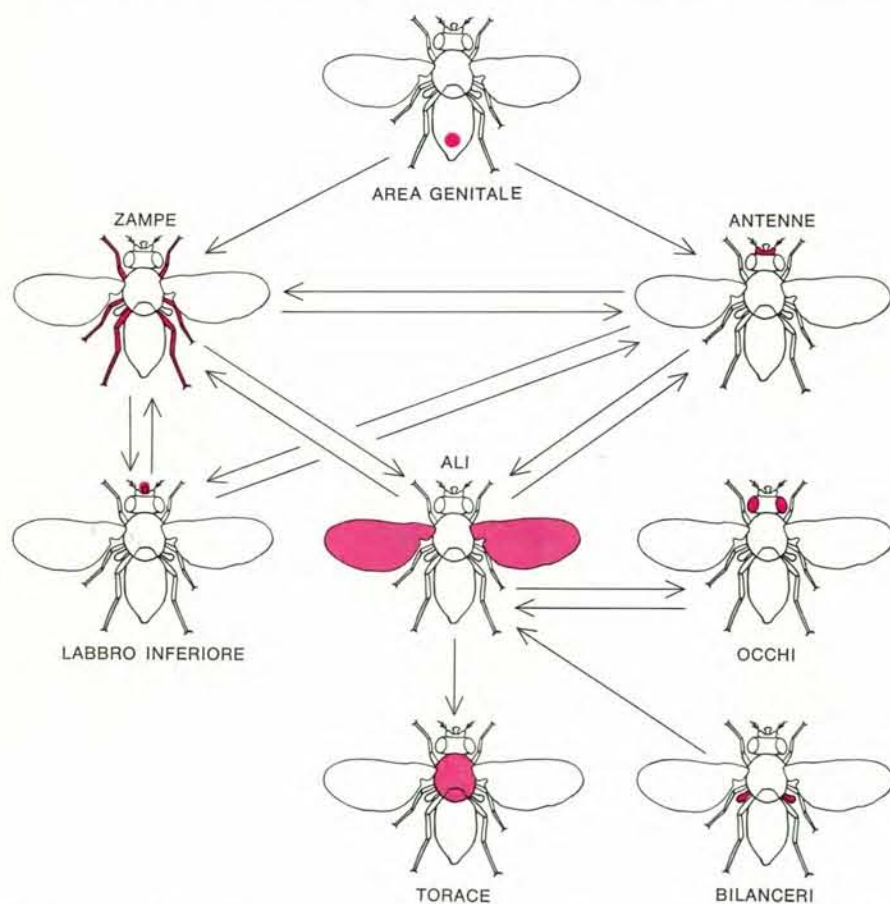
Il terzo risultato si ottenne quando parti delle colture vennero trapiantate di nuovo sulle larve. Tali trapianti si metamorfosarono insieme con il loro nuovo ospite. Con qualche eccezione, della quale riferirò piú oltre, le porzioni prese dalle colture permanenti si differenziarono normalmente in quelle strutture adulte per le quali erano determinate. La capacità di svilupparsi normalmente è perciò conservata nelle nostre colture; durante tutte le generazioni di trapianti il patrimonio genetico necessario per la differenziazione nell'adulto rimane intatto. Tale stabilità è insolita per cellule che crescono in colture permanenti. In tutte le cellule in divisione abbiamo sempre trovato corredi cromosomici normali.

I successivi trasferimenti di cellule da una singola porzione di disco imaginale sono qui raffigurati per 4 generazioni. Una porzione di disco genitale tolta da una larva è divisa fra un ospite adulto (*1, a sinistra*) e una larva (*a destra*); questa è portata alla maturazione in modo che si possa controllare che cosa avviene delle cellule trasferite. Il processo è ripetuto (2) ritrapiantando le cellule del disco dell'ospite « ceppo ». Dopo il successivo trapianto (*3*) le cellule del ceppo sono cresciute sufficientemente di numero da permettere due trapianti di controllo e due trapianti per mantenere il ceppo. Nella successiva generazione (*4*) i numeri si quadruplicano. Le cellule dei dischi sono vissute e si sono moltiplicate per piú di 150 generazioni di trapianti.

Ciò che ho detto fin qui è il prologo al capitolo piú importante della nostra storia. Immaginiamo ora di iniziare una coltura permanente con mezzo disco genitale di maschio. Una metà, la porzione detta « ceppo », in quanto serve a mantenere la coltura, viene trapiantata nell'addome di un moscerino adulto. L'altra metà, di controllo, viene trapiantata in una larva, all'interno della quale andrà incontro a differenziazione. Il « ceppo » cresce, dopo due settimane viene rimosso dall'ospite adulto e di nuovo viene poi diviso in due parti: una parte ceppo che conserverà la coltura, e una parte di controllo che mostra quale specie di differenziazione finale è capace di produrre. Dopo pochi trapianti le colture di solito crescono piú in fretta, cosicché diventa possibile far passare piú di un controllo attraverso la fase della metamorfosi e di trapiantare molte parti ceppo in modo da dare inizio a subcolture (*si veda l'illustrazione della pagina a fronte*).

Nel corso di 160 generazioni di trapianti per un periodo di 6 anni, sono state studiate parecchie migliaia di trapianti di controllo. In queste condizioni abbiamo avuto, indubbiamente, un'opportunità sperimentale unica. In ogni momento possiamo individuare di che tipo di differenziazione sono capaci le cellule in proliferazione.

Nelle prime generazioni di trapianto le parti di controllo del disco genitale si svilupparono in accordo con la loro originale determinazione. Cioè si differenziarono « autotipicamente » in organi genitali e organi anali. Dopo di che arrivò la sorpresa. Parti dei trapianti di controllo si svilupparono invece in organi della testa o in parti della zampa. In altre parole alcune popolazioni di cellule non si differenziarono piú secondo la linea di differenziazione delle loro progenitrici. Esse ora producevano organi « allotipici » che in un normale sviluppo si sarebbero originati solo da cellule del disco imaginale della testa o da quello della gamba. Per il cambiamento da un tipo di differenziazione a un altro abbiamo introdotto il termine di « transdeterminazione ».

L'intera storia di una coltura tipica può essere seguita nell'illustrazione di pagina 49. Se il lettore si soffermerà sull'illustrazione, vedrà che all'ottava generazione di trapianto (dopo circa quattro mesi) compaiono le strutture della testa e delle zampe. Una successiva transdeterminazione conduce, alla tredicesima generazione di trapianto, alla formazione di ali. Alla diciannovesi-

Gli organi del moscerino adulto si formano dalla maturazione di un certo numero di dischi imaginali presenti nel corpo della larva. Gli organi genitali e le porzioni dell'intestino e dell'addome si formano dalle cellule del disco genitale; la testa si sviluppa da 3 paia di dischi; ciascuna delle 6 zampe del moscerino si sviluppa da un disco a parte. Le cellule dei dischi che sono state mantenute in uno stato indifferenziato con successivi trapianti in ospiti adulti, continuano per anni a maturarsi in organi appropriati, quando si dà a esse la possibilità di compiere la metamorfosi. Dopo un certo numero di generazioni di trapianto, tuttavia, si verifica un cambiamento: le cellule dei dischi formano tessuti adulti di un altro tipo. La transdeterminazione generalmente segue la stessa sequenza (*si veda l'illustrazione qui sotto*) e il cambiamento è mantenuto ancora per generazioni.

La sequenza di transdeterminazioni di cellule di sette diversi tipi di dischi imaginali è indicata dalle frecce. Le cellule genitali, per esempio, possono cambiarsi in cellule della zampa o dell'antenna, mentre cellule della zampa e dell'antenna possono diventare cellule del labbro o delle ali. Nella maggior parte dei casi la transdeterminazione finale che risulta irreversibile è quella che porta da cellule delle ali a cellule del torace.

ma generazione di trapianto per la prima volta cellule del torace diventano stabili. Dopo molte colture di trapianto eravamo in grado di trarre alcune conclusioni generali sulla determinazione e la transdeterminazione. Dapprima abbiamo trovato che lo stato di determinazione può ripetersi senza alcun cambiamento quando le cellule vanno incontro a divisione; le caratteristiche di una determinata popolazione sono trasmesse per « eredità cellulare ». Ciò è vero per lo stato iniziale « autotipico » come per gli stati allotipici di nuova formazione, dopo che sono divenuti stabili per transdeterminazione. Per esempio, si può vedere nella stessa figura che gli organi genitali e anali compaiono nei trapianti di controllo senza interruzione dalla prima generazione di trapianto alla cinquantacinquesima. Ma anche popolazioni di cellule diventate recentemente allotipiche, si moltiplicano seguendo l'« eredità cellulare ». In certe linee, le strutture della testa, delle zampe, di parti del torace e delle ali continuano a comparire per anni.

Per ciascun particolare stato di determinazione esiste una probabilità o frequenza di transdeterminazione in una direzione specifica. Nelle cellule dei dischi genitali una transdeterminazione di primo ordine conduce pressappoco con la stessa frequenza alla formazione di cellule della testa o delle zampe. Da questi stati allotipici una transdeterminazione di secondo ordine porta a cellule delle ali. Non abbiamo mai osservato un cambiamento diretto da cellule genitali a cellule delle ali. Infine gli abbozzi del torace hanno origine da un'ulteriore transdeterminazione delle cellule delle ali.

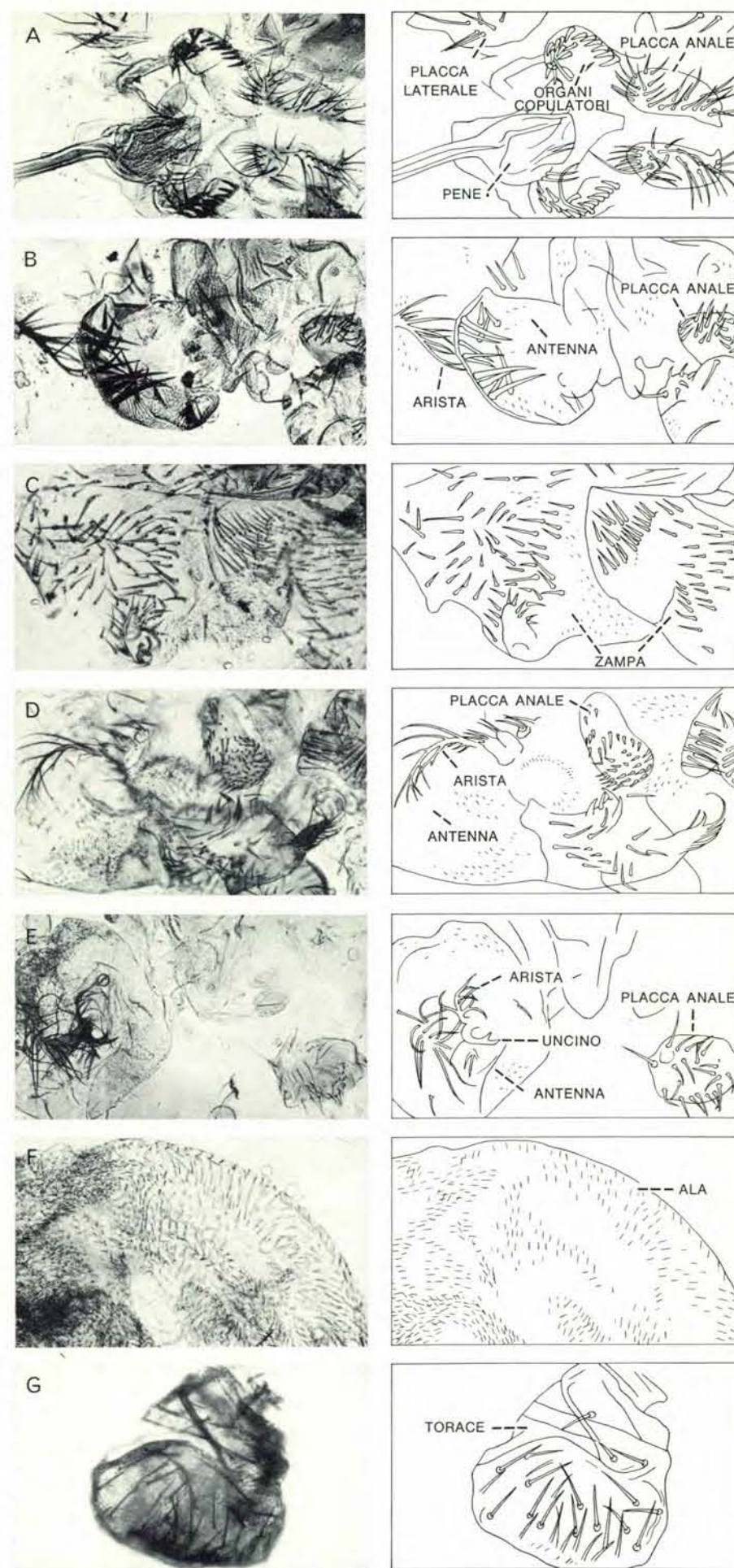
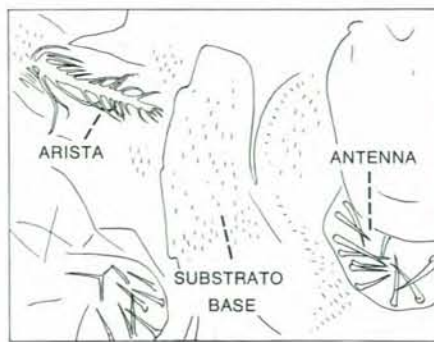
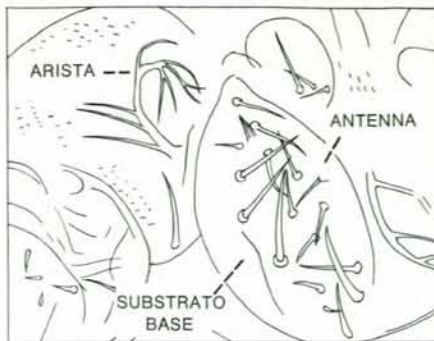
Queste sequenze sono riassunte nell'illustrazione qui a sinistra. Alcuni dei cambiamenti sono indicati come reversibili, altri come irreversibili. Per esempio, appare impossibile che cellule del torace diano origine a cellule genitali. La dinamica della transdeterminazione porta a una variazione nella frequenza di diversi tipi di differenziazione nei trapianti di controllo. Gli stadi iniziali rappresentati dalle strutture « autotipiche » diminuiscono o addirittura scompaiono, mentre gli stadi finali, come il torace, diventano sempre piú frequenti.

Ho portato come esempio gli esperimenti con i dischi genitali: però lo stesso comportamento viene osservato per gli altri dischi imaginali. Le cellule di questi dischi che sono inizialmente determinate per le parti della testa e della bocca, per parti delle ali o del torace,

subiscono anch'esse la transdeterminazione nel corso di un prolungato mantenimento in coltura. Ciascuna serie di cambiamenti ha la propria sequenza.

La transdeterminazione porta, come abbiamo visto, a un cambiamento nella eredità cellulare. Quali fattori causano questo cambiamento e su quali meccanismi è basato? Una risposta soddisfacente a queste difficili domande non può essere ancora data, ma possiamo avanzare alcune supposizioni. Un fatto stabilito dai miei collaboratori Tobler e Hansruedi Wildermuth è che la frequenza della transdeterminazione è in rapporto con la velocità di moltiplicazione delle cellule: piú in fretta esse si moltiplicano, piú alta è la probabilità di un cambiamento nell'eredità cellulare. Perciò interpretiamo la transdeterminazione come un fenomeno in qualche modo dovuto a cambiamenti nella dinamica di proliferazione, dalla quale potrebbe dipendere l'interazione funzionale dei diversi componenti cellulari.

La stabilità della determinazione è assicurata solo quando lo stesso gruppo di geni rimane attivo in successive generazioni di cellule. La transdeterminazione che porta a un nuovo tipo di eredità cellulare di lunga durata, invece, richiede l'azione di nuovi gruppi di geni. Dalle scoperte della genetica molecolare sappiamo che i geni sono attivati o repressi da particolari molecole. Quando le nostre cellule si dividono e crescono, queste molecole si diluiscono e se ne devono sintetizzare di nuove. Differenze e cambiamenti nelle velocità della divisione cellulare potrebbero, forse, in diverse popolazioni di una coltura, influenzare l'insieme dei componenti cellulari che controllano l'attività dei geni, e potrebbe risultarne un nuovo equilibrio fra i fattori interagenti. I membri

Le transdeterminazioni di cellule provenienti da un disco genitale avvenute durante 100 generazioni successive di trapianti, sono mostrate in una serie di microfotografie (*a sinistra nella figura, mentre a destra vi sono dei disegni nei quali vengono messi in evidenza gli organi che si sono formati prima e dopo la transdeterminazione*). Le generazioni sono quelle illustrate a pagina 49. Le strutture che si maturano dalla prima generazione di trapianto (*A*) sono tutte prodotti normali di cellule genitali. A partire dall'ottava generazione (*B*) alcune cellule genitali maturano come componenti della testa. Un campione della 24ª generazione (*C*) mostra strutture della zampa. Alla 37ª generazione (*D*) e alla 38ª (*E*) i campioni formano ancora organi genitali, ma anche strutture del capo. Le parti delle ali (*F*) compaiono alla 60ª generazione e le strutture toraciche (*G*), che sono i prodotti finali, compaiono invece dalla 100ª generazione in poi.

Strutture normali e anormali sono mostrate in 2 microfotografie e spiegate nei disegni che le accompagnano. In alto c'è un tessuto adulto della 31ª generazione di trapianto di cellule del disco genitale; esso si è sviluppato in elementi normali del capo. In basso vi sono gli stessi elementi, ma in forma anormale, che si ritrovano nel tessuto della 41ª generazione di trapianto. È stato dimostrato che queste anormalità sono ereditarie.

di un gruppo di geni precedentemente attivi che determinavano, per esempio, le zampe, potrebbero diventare inattivi. Il loro ruolo potrebbe essere preso, allora, dai geni determinanti per le ali.

Solo ulteriori esperimenti potranno portare a una migliore comprensione di tutto ciò che avviene nella transdeterminazione. Vorrei però richiamare l'attenzione su un altro fatto. Come abbiamo visto le cellule determinanti per gli organi genitali possono dar luogo a discendenti che si sviluppano alternativamente in un intero insieme di altre parti del corpo. Come il mio collaboratore Walter Gerhing ha dimostrato, sequenze di diverse differenziazioni possono insorgere in popolazioni di cellule (cloni), derivate da una singola cellula. Perciò la determinazione normale non è certamente basata su uno stato irreversibile delle unità cellulari. La capacità di intraprendere diverse vie di sviluppo è una caratteristica persistente delle cellule.

Fin qui ci siamo occupati di cambiamenti nelle colture che portano da un normale tipo di differenziazione a un altro tipo anch'esso normale. I trapianti allotipici del torace non possono, persino dopo sei anni di coltura, essere distinti da campioni di controllo autotipici, presi direttamente da un disco controllo di una larva. Oltre a tale differenziazione « normotipica », tuttavia, abbiamo incontrato in alcune linee cellulari un tipo « anormotipico » di sviluppo (*si veda l'illustrazione qui sopra*). In una linea che venne seguita dalla trentottesima generazione di trapianto alla novantesima, i piccoli peli e protuberanze (tricomi) che ricoprono la cuticola fra le setole appaiono ruvidi e un po' distorti. Analogamente l'arista, una struttura che sporge dall'antenna, è di forma anormale. D'altra parte le setole sono tutte normali. In un'altra linea le strutture di base sono del tutto normali, ma non si ha piú la formazione delle setole. Tali linee anormotipiche possono essere dovute a classiche mutazioni e non è mai stato osservato un ritorno alla normalità. Una volta che le caratteristiche anormali si sono stabilizzate, esse sono durate fino a quando abbiamo potuto osservare i discendenti. Un ulteriore tipo di sviluppo si è venuto a stabilire in altre linee. Queste colture hanno una velocità di accrescimento estremamente alta, ma i campioni di controllo o spariscono durante la metamorfosi, o passano attraverso lo stadio di pupa senza alcun cambiamento visibile. In ogni caso la capacità per una differenziazione adulta è perduta. Forse tali tipi di eredità cellulare sono dovuti a quel tipo particolare di mutazione che in un animale verrebbe chiamato fattore letale di sviluppo.

Tutti gli eventi messi in luce nelle colture di cellule derivate dai dischi imaginali dovrebbero ora venir paragonati a processi che avvengono in altri sistemi biologici. Qui ricorderò solo alcune somiglianze piú evidenti. La stabilità nell'eredità cellulare, cioè il tipo di stabilità osservato nelle nostre colture in moltiplicazione, si incontra, fra gli altri tessuti, nell'epidermide della pelle umana e nel midollo delle ossa umane. In questi tessuti le cellule del ceppo rimangono indifferenziate, ma si moltiplicano per tutta la vita dell'individuo mantenendo uno specifico stato di determinazione. In biologia sono noti altri fenomeni che sono analoghi alla transdeterminazione. Consideriamo certi tipi di rigenerazione. I biologi che lavorano sulla rigenerazione delle zampe degli anfibi hanno dimostrato che cellule precedentemente specializzate e differenziate generano cellule figlie che sono determinate per nuovi compiti. Negli insetti e nei crostacei si osserva occasionalmente che dal moncherino di un'antenna si può rigenerare una zampa o un occhio. Tale risultato allotipico è basato probabilmente sugli stessi eventi cellulari della transdeterminazione. Inoltre nella *Drosophila* parecchi mutanti del tipo detto omoiotico possono dar luogo a organi allotipici. Una di queste mutazioni è chiamata aristapedia: essa induce cellule dei dischi imaginali dell'antenna a differenziarsi, durante lo sviluppo, in parti della zampa anziché in strutture dell'antenna. L'effetto finale è lo stesso di quello che si verifica nelle nostre cellule in transdeterminazione, ma non si devono trascurare, però, alcune differenze essenziali tra i due casi. Gli effetti delle mutazioni omoiotiche sono dovuti a cambiamenti del materiale genetico, mentre nel caso della transdeterminazione noi ammettiamo che si tratti di un cambiamento nei sistemi di controllo che regolano l'attività dei geni che invece non sono mutati.

Poiché nelle colture si osservano tanti cambiamenti nel comportamento delle cellule che proliferano e nella loro eredità, si potrebbe essere tentati di paragonare alcuni dei nostri risultati con la proliferazione cancerogena e con la differenziazione aberrante nei tumori. Io credo che sia troppo presto per un confronto dettagliato. Tutto ciò che si può dire è che le cellule cancerogene, come le cellule transdeterminate, si discostano dal comportamento delle loro progenitrici andando incontro a nuovi tipi di sviluppo.

La combustione risonante nei razzi

I vari processi fisici e chimici che si svolgono durante la liberazione dell'energia in un motore a razzo possono produrre condizioni adatte a forti amplificazioni delle oscillazioni.

di J. George Sotter e Gary A. Flandro

I razzi a propellente solido furono probabilmente inventati in Cina circa mille anni fa. Essi non attrassero molto l'attenzione fino al XIII secolo d. C., quando si cominciò ad adoperarli in guerra. A partire dal quattordicesimo secolo circa rivaleggiarono costantemente con le armi da fuoco. Non molto tempo dopo il 1814, anno in cui Francis Scott Key parlò del « rosso bagliore dei razzi », però, il perfezionamento delle armi da fuoco sembrò aver destinato i razzi a un ruolo secondario. Perciò quando nel 1895 Pedro Paulet, un ingegnere peruviano, inventò il primo razzo a propellente liquido veramente efficiente, la cosa non destò eccessivo interesse.

Trent'anni piú tardi il fisico americano Robert H. Goddard lanciò un razzo a propellente liquido, azionato a ossigeno liquido e benzina. Goddard, il cui lavoro pose le basi della tecnologia dei propellenti liquidi, fu in quell'epoca considerato quasi un pazzo da alcuni degli specialisti americani che seguivano i suoi tentativi. I suoi risultati furono però presi in seria considerazione in Germania. Nel 1939 gli ingegneri missilistici tedeschi sollecitavano finanziamenti per la realizzazione di un grande razzo a propellente liquido, che poi divenne famoso come *Vergeltungswaffe* 2 (« arma della vendetta »).

La V2 inaugurò l'era spaziale e aprí la strada al viaggio dell'uomo verso la Luna. Tuttavia nel 1940 un gruppo inglese, sviluppando un piccolo razzo antiaereo a propellente solido, si imbatté nella combustione instabile, un fenomeno che da allora ha afflitto quasi tutti i grossi programmi di sviluppo dei motori a razzo. Durante una prova statica del razzo inglese al Woolwich Arsenal, a metà circa della combustione si manifestò prima un brusco aumento di pressione, fino a piú del doppio del valore previsto, e poi un rapido ritorno al valore normale. Un tale sbalzo avrebbe certamente distrutto un motore a razzo costruito per volare, ma gli esperimenti si svolgevano con una camera di prova molto piú resistente, che non ne fu danneggiata.

In seguito i tecnici inglesi, allo scopo di analizzare il fenomeno, realizzarono delle accensioni che potevano essere interrotte non appena appariva il salto di pressione. Esaminando la superfice del combustibile solido parzialmente bruciato essi notarono delle increspature. L'orientamento dei solchi suggeriva che all'interno del motore il gas era stato animato da un intenso moto vorticoso.

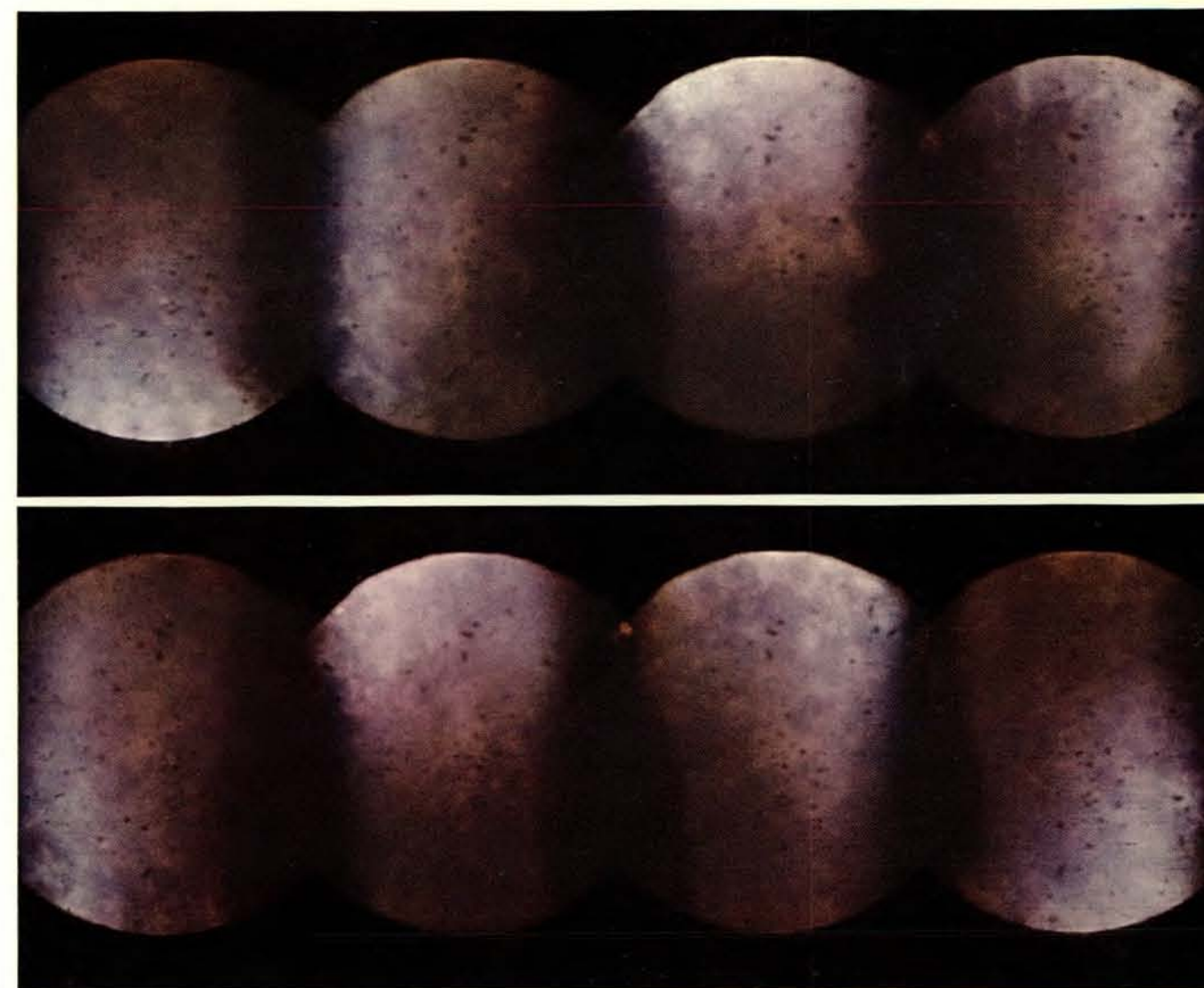
Evidentemente l'instabilità era imputabile a un moto irregolare del gas durante la combustione. I tecnici inglesi smorzarono questo moto in una maniera molto semplice: fissarono una riga di acciaio all'estremità anteriore della camera di combustione, in modo che sporgesse nella corrente gassosa. Tali « barre di risonanza », realizzate in forme e materiali diversi, sono tuttora adoperate per controllare la combustione instabile dei motori a propellente solido. Esse tuttavia presentano diversi inconvenienti. Al crescere delle dimensioni dei motori le barre devono essere sempre piú massicce; ostacolano l'accensione del propellente, riducono il carico utile e spesso non riducono sensibilmente l'instabilità. Per quanto si possa ricorrere talvolta a vari altri rimedi, la combustione instabile nei razzi a propellente solido rimane un problema.

I razzi a propellente liquido non mostrarono inizialmente fenomeni di combustione instabile. Il problema si presentò quando i progettisti del razzo tedesco V2 iniziarono la realizzazione di un iniettore di fabbricazione piú semplice rispetto a quello usato. Il loro nuovo iniettore funzionò bene in un piccolo motore di prova, ma, quando fu provato in un modello in scala naturale, il motore emise un chiaro ronzio accompagnato da forti vibrazioni. I tecnici tedeschi risolsero il problema semplicemente tornando all'iniettore originale. D'altra parte molti razzi moderni a propellente liquido sono sensibilmente diversi dalle V2; adoperano propellenti piú energetici e hanno iniettori che consentono combustioni piú efficienti. Tutti questi fattori sembrano favorire l'instabilità, sicché anche nei razzi a propellente liquido l'instabilità rimane un problema da risolvere.

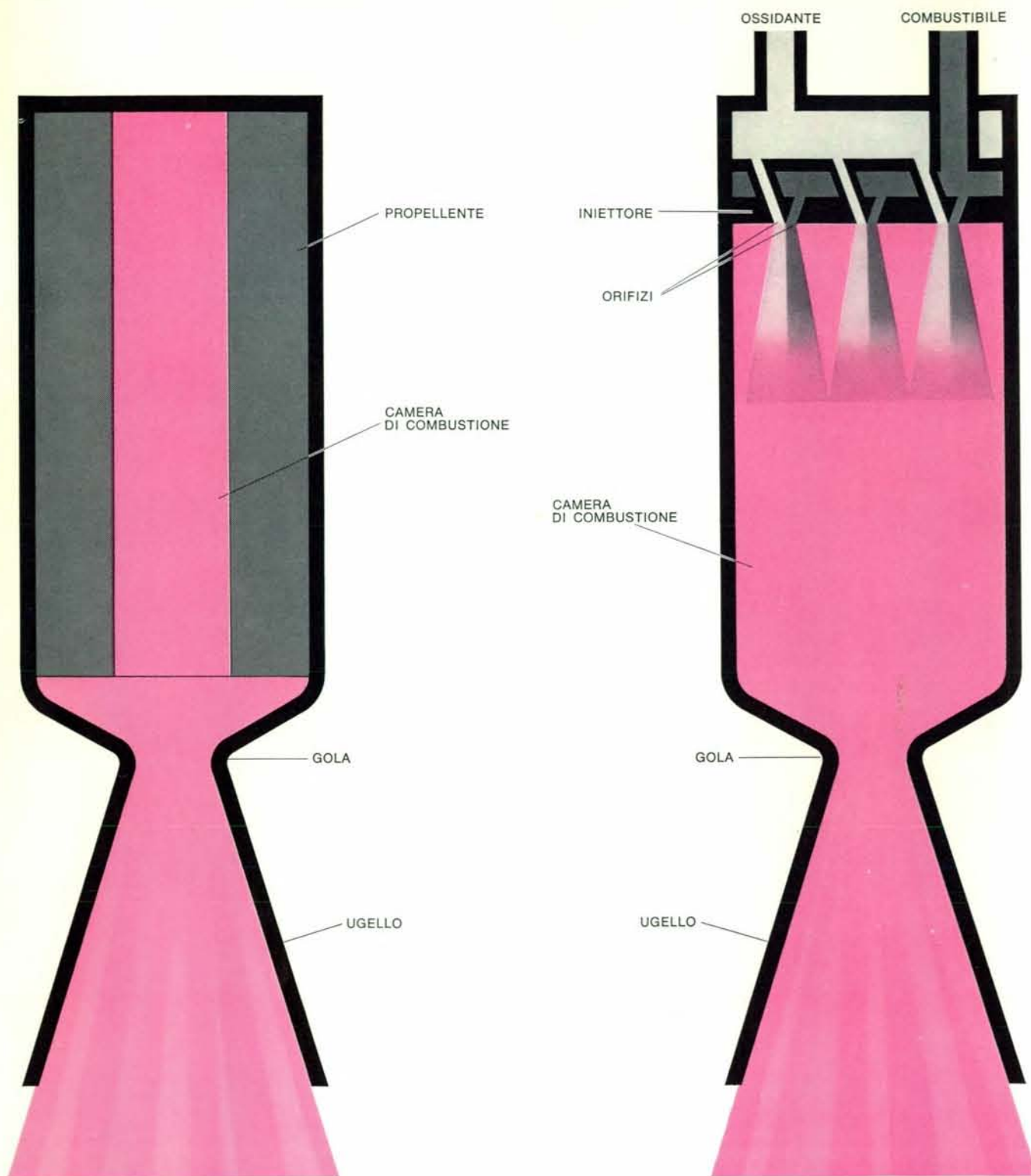
Per definizione, il razzo è un veicolo capace di propellersi senza ricorrere all'ambiente in cui si muove. I motori d'automobile e i motori a getto con turbina a gas impiegano l'aria dell'ambiente come ossidante per la combustione di un combustibile. Il razzo chimico trasporta con sé sia l'ossidante sia il combustibile, e quindi può funzionare anche dove non esiste l'aria, per esempio nello spazio interstellare.

La maggior parte dei razzi oggi in uso sono razzi chimici e funzionano mediante liberazione dell'energia chimica del combustibile e dell'ossidante (che insieme sono chiamati *propellenti*). Esistono o sono in fase di realizzazione anche pochi altri tipi di motore a razzo, come il motore nucleare e quello elettrico, che però, per le loro caratteristiche particolari, non saranno trattati in questo articolo.

Un razzo chimico può essere a propellente solido o liquido: in entrambi i casi la combustione trasforma i propellenti in gas caldi. I gas si espandono, creano un'alta pressione che li costringe a fluire attraverso un ugello dove si raffreddano mentre la loro energia termica si trasforma in energia cinetica. I gas vengono scaricati dall'ugello a ve-

Questa vista interna di un razzo a propellente solido mostra un vortice prodotto da instabilità di combustione. Le immagini, tratte da una ripresa cinematografica realizzata da uno degli autori (Sotter) e da J. Swithenbank all'Università di Sheffield, sono state riprese attraverso una finestra nella parte anteriore del razzo e verso l'ugello attraverso cui il gas si scarica.

Onde di detonazione in un razzo a propellente liquido; in questa sequenza di fotogrammi, tratti da un film ottenuto da Richard M. Clayton al Jet Propulsion Laboratory, le onde si muovono in senso orario. La cinepresa, montata in un alloggiamento protettivo, era sistemata nel gas di scarico appena all'esterno dell'ugello. La vista è quindi rivolta verso l'interno del razzo in direzione dell'iniettore, che è costituito da una piastra perforata di metallo disposta nella testata del motore. Il combustibile e l'ossidante vengono introdotti attraverso l'iniettore nella camera di combustione, dove si mescolano e bruciano.

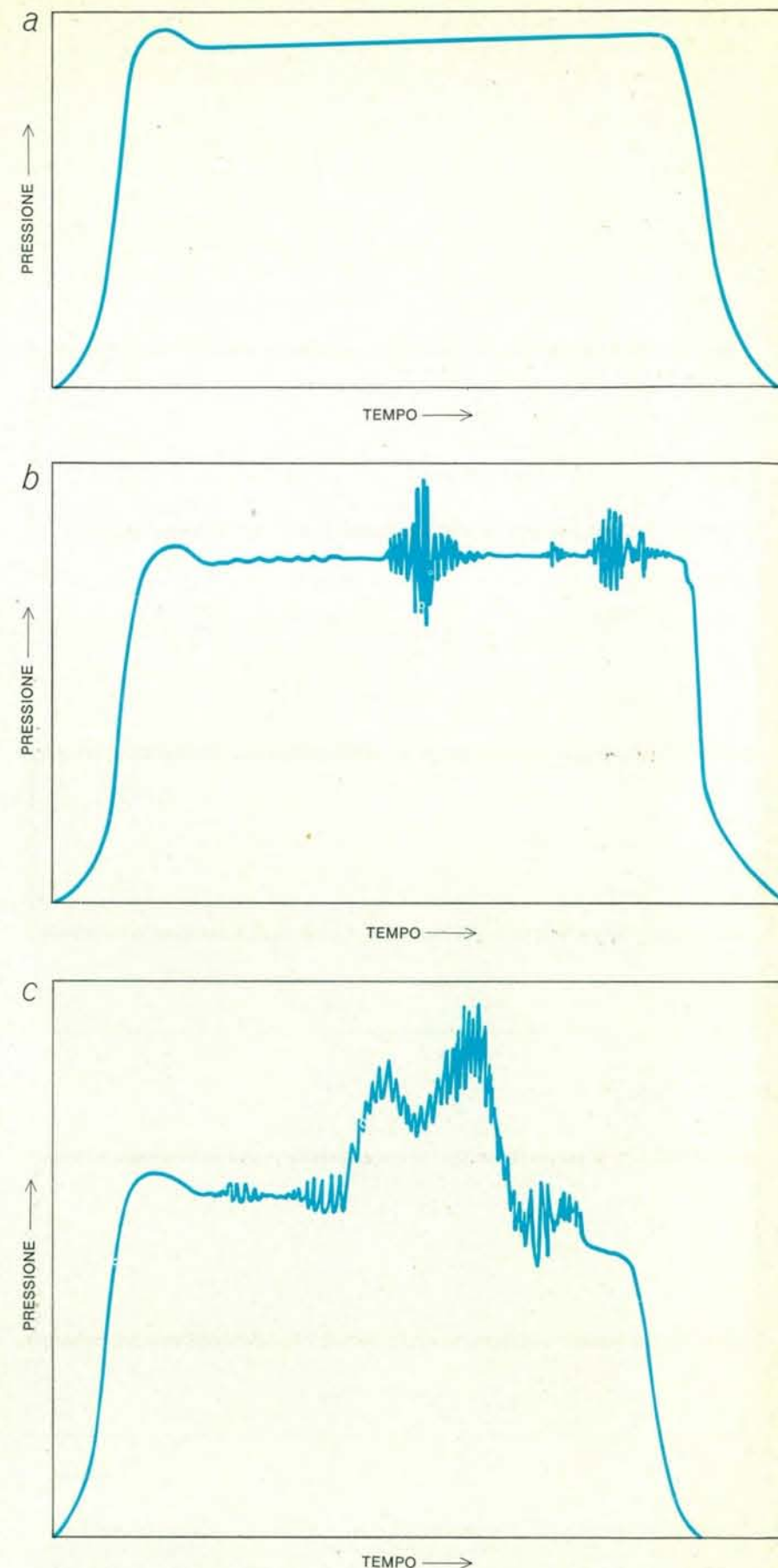
locità supersonica, e il movimento di avanzamento del razzo è effetto della reazione che corrisponde alla loro espulsione.

Un propellente solido può essere realizzato mescolando un combustibile con un ossidante (all'atto del mescolamento uno degli ingredienti è in ogni caso allo stato liquido, ma successivamente rapprende allo stato solido). Una miscela tipica è il perclorato di ammonio in una matrice di combustibile organico, per esempio gomma. Si può anche usare una miscela omogenea di sostanze autoossidanti, come la nitrocellulosa e la nitroglicerina. In entrambi i casi il propellente è fuso in una forma che presenta una cavità centrale per tutta la sua lunghezza. La combustione si sviluppa solo sulla superficie libera del materiale.

I razzi chimici, che costituiscono attualmente il principale tipo adoperato, possono far uso di propellenti sia solidi (*a sinistra*) sia liquidi (*a destra*). Un razzo a propellente solido può essere considerato il più semplice dei sistemi propulsivi, in quanto consiste essenzialmente in un ugello e in una camera contenente la carica. Un razzo a propellente liquido brucia un combustibile e un ossidante, i quali sono conservati separatamente nel veicolo e vengono introdotti nella camera di combustione attraverso una piastra metallica perforata. Una tipica combinazione di propellente è costituita da cherosene e ossigeno liquido.

In un razzo a propellente liquido l'ossidante e il combustibile sono introdotti separatamente dai serbatoi di rifornimento nella camera di combustione. Passano attraverso iniettori che li atomizzano e ne provocano il mescolamento nelle giuste proporzioni in modo che possano rapidamente vaporizzare e bruciare. L'ossigeno e il fluoro liquidi sono alcuni degli ossidanti usati, mentre come combustibili si preferiscono il cherosene, l'alcool e l'idrazina.

Un motore a propellente solido non richiede nessuna pompa né sistema di pressurizzazione, e quindi è meccanicamente piú semplice di un motore a propellente liquido. D'altra parte, un motore a propellente solido non è controllabile come uno a liquido. Per esempio, solo recentemente è stata resa possibile l'interruzione di funzionamento di un motore a propellente solido e la sua riaccensione, mentre la cosa è facilmente realizzabile in un motore a propellente liquido. In linea di massima, un motore a propellente liquido è adatto a fornire prestazioni piú elevate e piú elastiche di uno a propellente solido, e ciò spiega perché la massima parte dei motori per veicoli spaziali sia a propellente liquido.

I propellenti, sia in forma solida sia liquida, possono essere visti come un grande serbatoio di energia che può essere spillata tramite i diversi processi chimico-fisici che si svolgono durante l'operazione e che danno luogo alla possibilità di oscillazioni autoeccitate nel sistema. Il comportamento oscillatorio è sempre possibile quando una sorgente di energia si trova in presenza di un corpo oppure di un ambiente capace di moto vibratorio e quando fra la sorgente di energia e il sistema vibrante si realizza un circuito di retroazione. Il piú noto esempio di questo fenomeno è probabilmente rappresentato dalla risonanza che si verifica talvolta quando un suono casuale viene raccolto da un microfono e attraverso l'amplificatore perviene all'altoparlante, che lo emette in modo che venga nuovamente raccolto dal microfono. L'impulso acquista intensità ogni volta che passa attraverso l'amplificatore (che è una sorgente di energia), e, poiché il processo si ripete piú volte, il suono può intensificarsi fino a divenire un fischio lacerante. In pratica si raggiunge poi un equilibrio fra quello che si guadagna e quello che si perde, in quanto il volume del suono viene livellato dai fenomeni di smorzamento (provocati dall'acustica delle pareti, dalle caratteristiche dell'amplificatore e da altre condizioni).

Un fenomeno analogo si verifica durante la combustione instabile dei raz-

Tracce di registrazioni di pressione, che caratterizzano diversi tipi di combustione nei razzi. La combustione stazionaria normale (*a*) provoca una curva con profilo continuo. La combustione oscillatoria (*b*) rivela diverse frequenze, a volte con ampiezze variabili. L'allontanamento radicale della pressione dal valore normale (*c*) viene chiamato combustione irregolare; questa sembra essere innescata dalla combustione oscillatoria.

zi; in tale analogia la regione di combustione della camera di combustione corrisponde all'amplificatore. In certe condizioni questa regione può amplificare eventuali onde soniche, e, se le pareti della camera forniscono un'opportuna riflessione, l'intensità delle onde può crescere fino a valori che dipendono soprattutto dall'entità dello smorzamento in atto.

Il comportamento oscillatorio nei motori a razzo si manifesta in diverse forme, alcune delle quali sono ancora poco note dal punto di vista teorico. Una di queste forme, che si verifica nei motori a propellente liquido, è una vibrazione di bassa frequenza, che si rivela con uno scoppiettio causato dall'accoppiamento fra il sistema di alimentazione del propellente e il processo di combustione. Tale accoppiamento deriva dal fatto che la portata di propellente affluente è sensibile alla pressione esistente nella camera di combustione.

Non appena la pressione nella camera cresce, inizialmente forse per un disturbo casuale, la portata di propellente attraverso l'iniettore tende a diminuire. La riduzione del flusso provoca una diminuzione di pressione nella camera di combustione che favorisce un aumento del flusso di propellente. Questo tipo di instabilità può essere evitato se i tempi caratteristici di incremento e decremento della pressione nella camera e delle portate di propellente sono notevolmente diversi. Il problema può essere risolto mediante rivestimento delle linee di alimentazione del propellente, o cambiando il disegno dell'iniettore oppure con altri criteri che abbiano lo stesso effetto: realizzare cioè tempi caratteristici cosí differenti che la risonanza non possa verificarsi.

Una forma di instabilità piú distruttiva, presente sia nei razzi a propellente liquido sia in quelli a propellente solido, è quella chiamata « instabilità acustica ». Sostanzialmente è caratterizzata da variazioni ad alta frequenza della pressione all'interno della camera; il motore emette in questo caso un sibilo che in molti casi può essere udito distintamente.

L'instabilità acustica può avviarsi con piccoli disturbi nel gas in camera di combustione; questi disturbi vengono poi amplificati dal processo di combustione e dal flusso dei gas nel motore. Le onde soniche si riflettono alle pareti della camera venendo cosí nuovamente immesse nel sistema. Nei motori a propellente solido il fenomeno conduce a bruschi aumenti di pressione; nei razzi a propellente liquido sono le stesse oscillazioni che producono gravi danni in quanto causano forti accelerazioni vibratorie e determinano un eccessivo scambio termico con le pareti della macchina.

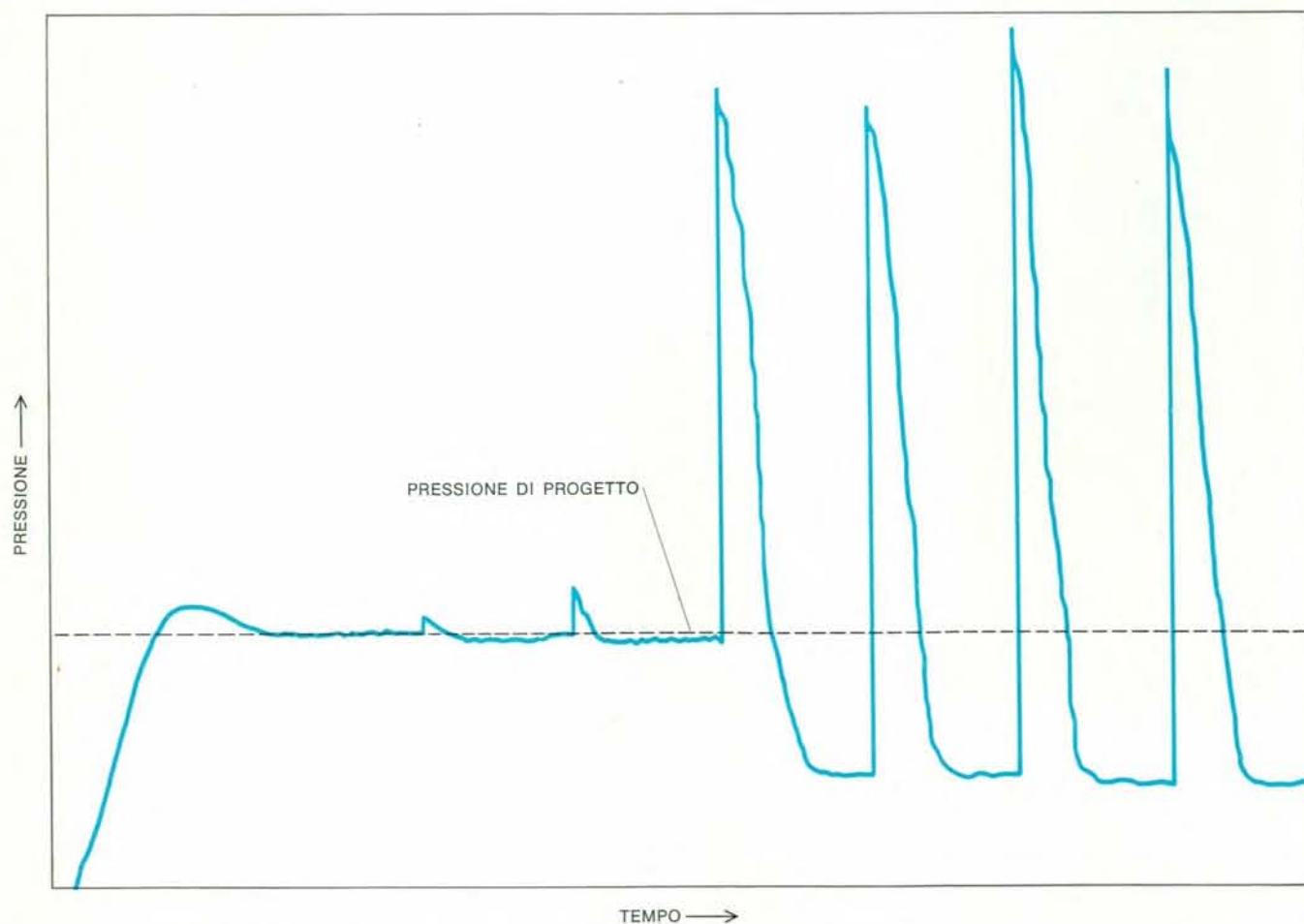
Per eliminare la combustione oscillatoria nei razzi si è proceduto essenzialmente per tentativi, il che è inevitabilmente costoso e richiede notevole tempo. Le difficoltà sarebbero state ridotte se fosse esistita una teoria capace di predire il verificarsi di onde di grande ampiezza in un motore di date dimensioni e di permettere l'indagine anticipata dei possibili sistemi per eliminarla. Il primo tentativo in tal senso fu pubblicato nel 1942 dal fisico russo Ya. B. Zel'dovich, ma il tentativo di far uso della sua analisi per predire la stabilità di motori reali non ebbe esito. Nonostante gli intensi sforzi compiuti in seguito da diversi ricercatori, finora non esiste ancora una teoria completamente soddisfacente. Le ragioni di tale situazione saranno forse evidenti se chiariremo brevemente i processi fisici che si verificano durante una combustione oscillatoria.

L'acustica ha fornito una struttura analitica di riferimento per molti dei lavori teorici sulla instabilità di combustione nei razzi. Una formulazione matematica particolarmente utile è costituita dall'equazione delle onde soniche, che definisce i diversi modi di oscillazione di un gas. In un cilindro circolare chiuso, geometria a cui può assimilarsi quella di certi motori a razzo, sono possibili tre modi di oscillazione « pura »: longitudinale, radiale e tangenziale. I modi longitudinale e radiale, come il loro nome indica, implicano esclusivamente movimenti lungo l'asse o il raggio del cilindro. Il modo tangenziale implica movimenti che hanno ovunque componenti sia tangenziali sia radiali, eccetto che alle pareti del cilindro, dove sono puramente tangenziali. Ovviamente i tre modi possono verificarsi simultaneamente e in diverse combinazioni; per di piú sono possibili vari modi le cui componenti longitudinale, radiale e/o tangenziale sono combinate intrinsecamente.

L'equazione delle onde soniche è abbastanza facile da risolvere per camere di geometria semplice, principalmente perché si tratta di un'equazione lineare. Questo importante aspetto è però ottenuto a spese del fatto di supporre che le variazioni di pressione (e le corrispondenti variazioni di velocità del gas) siano sempre infinitesime. Nell'acustica classica tali assunzioni corrispondono bene alla realtà: per esempio una comune conversazione provoca variazioni di pressione dell'ordine di sette centomilionesimi di chilogrammo al centimetro quadrato. In un motore a razzo invece sono state osservate oscillazioni di pressione fino a un miliardo di volte piú forti delle precedenti (cioè pari a 70 chilogrammi al centimetro quadrato). Perciò, come si può immaginare, esistono fondamentali differenze fra i fenomeni acustici di un motore a razzo e quelli di una camera chiusa risonante; le forti oscillazioni che si verificano in

In questa figura è rappresentata schematicamente un'oscillazione acustica di tipo longitudinale in un tubo chiuso. L'oscillazione acustica, o onda sonica, si immagina presente all'inizio dell'osservazione del processo (*a*), istante in cui la distribuzione di particelle, o gruppi di molecole, è quella rappresentata dalla concentrazione dei punti. Il grafico inferiore rappresenta la distribuzione di pressione. Subito dopo le particelle vengono uniformemente disperse (*b*), ma la velocità di molte di esse è rivolta verso l'interno, come indicato dalle frecce. Le particelle si addensano poi al centro (*c*), innalzando la pressione che le costringe ad arrestarsi e poi le spinge di nuovo verso le estremità.

La forma tangenziale di instabilità si manifesta in due modi: l'onda stazionaria (*a sinistra*) e quella viaggiante (*a destra*). In entrambi i casi in figura sono rappresentate per un mezzo ciclo le variazioni di pressione, dai valori alti (*zone scure*) a quelli bassi (*zone chiare*), e quelle di velocità (*frecce*). Il modo stazionario è analogo allo sciacquio che si produrrebbe inclinando una tazza di caffè avanti e indietro, il modo viaggiante si verifica facendo ruotare la tazza circolarmente.

Un'onda rotante assomiglia a una detonazione, e presumibilmente assume tale aspetto nel viaggiare intorno alla camera di combustione di un razzo a propellente liquido. Nella zona luminosa antistante l'iniettore (*a sinistra*) l'onda è ripida e molto forte. Il movimento avviene in senso orario visto dallo scarico dell'ugello. Una formazione simile è mostrata a pag. 59 in basso.

La traccia di registrazione di pressione di un'onda rotante è nettamente diversa dalla forma sinusoidale prodotta da un disturbo acustico. L'onda può svilupparsi da un disturbo acustico o può apparire bruscamente, come nel caso riportato in questa figura. Attraverso un'onda di questo tipo la pressione può raggiungere valori anche venti volte superiori a quelli di progetto.

un motore a razzo non possono quindi essere descritte dall'equazione classica delle onde.

Oltre agli effetti delle grandi ampiezze, l'acustica dei motori a razzo è complicata dalla presenza dei solidi e dei liquidi in combustione: questi infatti possono fornire enormi quantità di energia acustica al sistema. Un'altra sorgente diretta di energia acustica è rappresentata dal flusso ad alta velocità dei gas attraverso la camera di combustione. È necessario quindi ricavare equazioni piú complesse, partendo dalle leggi di conservazione della massa, dell'energia e della quantità di moto.

In un gas le onde soniche si propagano per effetto delle collisioni molecolari, e la velocità del suono è dello stesso ordine di grandezza di quella media molecolare. Tuttavia in acustica non è in genere necessario far riferimento alla distribuzione statistica delle molecole, il cui numero è estremamente alto. Il gas può essere immaginato come composto di « particelle », ossia gruppi di molecole, ciascuna delle quali comprende un alto numero di molecole. Prese assieme le particelle costituiscono un « continuo » avente certi valori di pressione, densità, temperatura e velocità in ogni punto. Per la fase gassosa queste quattro variabili sono legate alle leggi della conservazione della massa, dell'energia e della quantità di moto, nonché dall'equazione di stato (per esempio quella del gas ideale). Analoghe relazioni valgono per le fasi liquide o solide del propellente.

I due gruppi di equazioni sono strettamente connessi, in quanto il comportamento dei gas influenza quello della fase condensata e viceversa. Il calore proveniente dai gas combusti, che hanno una temperatura tipica sui 2700 gradi centigradi, provoca l'evaporazione dei propellenti, i quali, al di là di un sottile strato alla superficie riscaldata, possono trovarsi alla temperatura normale della camera o al di sotto di essa. I gas in rapido movimento provocano un forte aumento dello scambio termico e possono erodere i solidi o frantumare le goccioline di liquido, in maniera che momentaneamente l'evaporazione può venire fortemente intensificata. Dopo breve tempo i vapori addizionali liberano tramite la combustione la loro energia chimica. Se solo una piccola percentuale di questa energia viene rivolta all'effetto di eccitazione delle oscillazioni, possono nascere onde di pressione di enorme ampiezza.

Nello studio della combustione instabile dei razzi, l'interesse è rivolto solo alla parte del motore che precede la gola dell'ugello. Nella parte divergente che segue la gola il flusso è supersonico; ciò vuol dire che qualunque piccolo disturbo di pressione (il quale viaggia alla velocità del suono) sarà espulso così rapidamente verso lo scarico da non poter in alcun modo raggiungere la camera di combustione. Per tale ragione la camera di combustione viene considerata limitata verso lo scarico dalla « superficie sonica » gassosa che si verifica nella gola.

I modi tangenziali di oscillazione sono quelli di maggiore interesse perché sembrano costituire il tipo di instabilità piú distruttiva. La migliore definizione del primo modo tangenziale è quella di « sguazzamento » gassoso intorno alla camera di combustione; esso rassomiglia infatti allo sguazzamento di un liquido in un recipiente cilindrico. I nodi di velocità, che sono punti in cui la velocità del gas si annulla, possono restare stazionari alle pareti della camera. In tal caso il moto viene detto onda stabile o di « scuotimento ». Altrimenti i nodi possono spostarsi lungo le pareti della camera, e in questo caso il moto viene definito come forma viaggiante o « sguazzante » di instabilità tangenziale. La forma viaggiante, quando si verifica con forti ampiezze, produce numerosi effetti degni di nota. Uno di questi è la formazione di onde di detonazione, le quali possono essere impropriamente definite onde d'urto eccitate dalla combustione (*si veda la figura in alto nella pagina a fronte*). Un'onda d'urto (un ben noto esempio di tale fenomeno è rappresentato dal *bang* sonico degli aerei supersonici) consiste in un disturbo frontale estremamente ripido che a volte si sviluppa da disturbi di tipo sinusoidale, ossia acustici. Poiché i gas si riscaldano per effetto della compressione, la parte ad alta pressione dell'onda sinusoidale (dove la velocità del suono, che è proporzionale alla radice quadrata della temperatura assoluta, può essere considerevolmente piú alta di quella media) può viaggiare piú rapidamente della parte che si trova a pressione minore. Di conseguenza le onde si innalzano, piú o meno come avviene per le onde del mare che formano frangenti in prossimità della spiaggia.

Una volta che in un motore a propellente liquido si sia formata un'onda d'urto, i forti gradienti di temperatura, pressione e velocità possono provocare drastiche variazioni dei gradienti di combustione. Alle pareti della camera l'onda d'urto può accrescere lo scambio termico fino a valori superiori a quelli ottenuti in una fiamma ossidrica. La camera ne può venire perforata in meno di un secondo.

In un motore a propellente solido la parete è protetta dal propellente stesso. In questo caso l'effetto piú notevole dell'onda tangenziale viaggiante è costituito dal brusco incremento che essa produce nelle pressioni operative medie del motore. L'effetto può essere tale da provocare l'esplosione del motore a razzo, o nel migliore dei casi ridurne le prestazioni.

Tale fenomeno viene chiamato combustione irregolare. Fotografie riprese attraverso la testata di un motore a propellente solido suggeriscono che l'irregolarità è il risultato della formazione di un vortice – ossia di un moto gassoso analogo a quello che si ha in un ciclone (*si veda la figura in alto a pag. 59*). Il vortice viene creato dall'onda tangenziale viaggiante; all'interno del vortice esiste un nucleo a bassa pressione analogo all'occhio di un ciclone, e l'effetto combinato del nucleo a bassa pressione e della vorticosità può ostacolare fortemente il flusso di gas attraverso l'ugello. In tal caso può avvenire che i gas vengano prodotti all'interno del motore piú rapidamente di quanto non siano scaricati attraverso l'ugello: ne risulta un aumento di pressione nel motore. Con alcuni propellenti il gradiente di combustione viene innalzato dall'aumento di pressione e dall'alta velocità esistente nel vortice, sicché la pressione all'interno della camera sale sempre di piú. Di conseguenza, queste condizioni possono provocare l'esplosione del motore.

Quando il vortice è meno intenso, il suo effetto piú evidente consiste nel produrre una coppia di torsione attorno all'asse longitudinale del razzo. La coppia è di difficile rilevamento nelle prove a terra, e è stata quindi raramente misurata; l'esistenza del vortice e della coppia associati è stata rivelata solo talvolta in prove di volo.

Furono appunto le manifestazioni della formazione di un vortice che il gruppo inglese scoprí nel 1940 al Woolwich Arsenal. Per ragioni difficili da spiegare, fin da quell'epoca nello studio dei razzi a propellente solido sembrano essere stati ampiamente ignorati sia l'esistenza di vortici sia i possibili effetti connessi a questo fenomeno. Solo negli ultimi anni si è compresa l'importanza di questo processo.

Un sistema di smorzamento in un razzo a propellente solido può essere realizzato con diaframmi disposti fra segmenti di propellente. In figura è rappresentata in sezione una tipica configurazione di propellente: tale forma ha lo scopo di disporre di una grande superficie esterna, dove la combustione si svolge. I diaframmi hanno lo scopo di bloccare le instabilità di combustione prima che assumano proporzioni significative.

progettisti che affrontano il problema di eliminare l'instabilità in un razzo si sforzano in genere di aumentare l'influenza degli effetti smorzanti. Una tecnica che viene alle volte impiegata con i razzi a propellente liquido consiste nel disporre all'interno della camera di combustione delle pareti perforate, acusticamente assorbenti. Tale soluzione non è valida per i motori a propellente solido, dove il propellente stesso riveste le pareti. È stato trovato utile in tal caso aggiungere al propellente solido delle particelle metalliche; in questo caso i prodotti gassosi della combustione contengono particelle solide o liquide di ossido metallico in sospensione. Le particelle di ossido dissipano energia acustica: l'effetto è abbastanza simile all'attenuazione del suono che si verifica nella nebbia, e in molti motori a propellente solido il sistema elimina l'instabilità di combustione.

Dove queste varie tecniche si mostrano inadeguate o inaccettabili esistono altre alternative. Per un razzo a propellente solido è possibile cambiare la composizione o la geometria del propellente. Nei razzi a propellente liquido si può cambiare il rapporto combustibile/ossidante oppure il disegno dell'iniettore.

Queste tecniche possono comportare una perdita nella efficienza di combustione. Dato che i motori a razzo sono molto efficienti nella utilizzazione dell'energia disponibile, potrebbe sembrare che una tale perdita sia un prezzo abbastanza piccolo da pagare in cambio di una combustione stabile. Tuttavia con un veicolo delle dimensioni del razzo americano Saturno V una perdita del 3 % nella efficienza di combustione richiederebbe l'uso di un altro serbatoio pieno di propellente per realizzare la stessa missione, e il risultato sarebbe ovviamente una drastica riduzione del carico utile del veicolo.

Una ricorrente difficoltà per i progettisti è costituita dal fatto che in molti motori instabili è difficile riprodurre i risultati delle prove. Per i razzi a propellente liquido la difficoltà si verifica nell'idraulica del sistema di iniezione. Un iniettore tipico consiste essenzialmente in una piastra di metallo spessa circa sei millimetri e con dei fori attraverso i quali il propellente viene introdotto nella camera di combustione. La presenza, all'estremità superiore dei fori, di bordi finemente scabrosi (spesso in associazione con forti velocità trasversali nel flusso di propellente), rende imprevedibile il flusso del liquido attraverso i singoli fori. Inoltre piccole variazioni da motore a motore nella realizzazione dei fori degli iniettori rendono poco probabile che due motori, apparentemente identici e usciti dalla stessa linea di produzione, abbiano poi un'identica stabilità.

Problemi analoghi si incontrano nei motori a propellente solido. Per esempio, possono avere un notevole effetto sulla stabilità variazioni nella temperatura del propellente. Può allora accadere che un motore, perfettamente funzionante in una giornata fredda, presenti una forte instabilità in una giornata calda.

La maggiore speranza per il progresso nella prevenzione dell'instabilità di combustione dei propellenti solidi è probabilmente costituita da un sistema di laboratorio chiamato bruciatore a T: si tratta di un motore a propellente solido di forma a T, che viene usato per misurare l'attitudine all'amplificazione nei vari propellenti. Sia la teoria sia gli esperimenti hanno mostrato infatti che la maggior parte dei propellenti è capace di amplificare oscillazioni in un largo campo di frequenze. Il bruciatore a T fornisce un ambiente controllato nel quale i meccanismi di smorzamento e la geometria della camera di combustione sono relativamente semplici. Il bruciatore a T, essendo adatto a misurare le amplificazioni, aggiunge un importante tipo di informazione alle cognizioni dello studioso, in quanto gli rende possibile caratterizzare l'aspetto meno noto dei sistemi instabili: il processo stesso cioè di combustione. Si può sperare che la conoscenza delle caratteristiche di amplificazione di un propellente, ottenuto molto economicamente con il bruciatore a T, potrà essere eventualmente applicata in modo attendibile alla previsione di ciò che accade in un grande motore a razzo.

Dato che l'instabilità di combustione è sensibile a piccole variazioni della geometria della macchina e delle condizioni operative, un particolare motore deve essere sottoposto a un gran numero di accensioni prima che i suoi progettisti possano dire con sicurezza che esso sia privo di instabilità. In un grande motore queste prove incidono fortemente sul costo di realizzazione e è quindi importante ideare adatte teorie sull'instabilità e criteri economici di prova delle caratteristiche acustiche del propellente. Fino a quando l'instabilità di combustione non sarà nota al punto da poter essere eliminata quando il motore è ancora nella fase di progetto, si dovrà continuare a eseguire lunghe prove di stabilità sui razzi, in particolare nei casi in cui la vita degli astronauti potrà dipendere dal funzionamento sicuro e fidato del motore.

L'espansione dei fondi oceanici

I fenomeni geofisici, dai terremoti alla deriva dei continenti, stanno forse per essere spiegati con una nuova teoria che promette di correlare il magnetismo terrestre con le dinamiche interna e orbitale della Terra.

di J.R. Heirtzler

In ogni campo della scienza sono rare le nuove teorie che razionalizzino globalmente un gran numero di osservazioni e che spieghino tutti i principali aspetti del mondo fisico. Una sintesi di questo genere sembra attualmente a portata di mano in geofisica. Negli ultimi anni si è affacciata una nuova teoria sui movimenti dei fondi oceanici. Essa riguarda immense e un tempo insospettate forze che rimescolano l'interno della Terra, determinando la disposizione dei bacini oceanici e delle masse continentali quali sono oggi. La teoria è basata su una grande varietà di osservazioni e di ipotesi che riguardano la morfologia dei fondali oceanici e la distribuzione dei sedimenti, le faglie e i terremoti, la struttura interna del globo, il campo magnetico terrestre e le sue periodiche inversioni. Questa teoria è inoltre perfettamente compatibile con quella della deriva dei continenti. Le due teorie assieme sono già riuscite a spiegare molte strutture della superficie terrestre e a chiarire la natura dei processi che si svolgono all'interno della Terra. È anche possibile che la loro importanza non sia stata ancora compiutamente apprezzata: esse potrebbero condurre infatti a una sintesi globale che correli il magnetismo terrestre con le dinamiche interna e orbitale della Terra.

Sviluppo della teoria

Le premesse per la scoperta della espansione dei fondi oceanici furono poste dal lungo dibattito sulla deriva dei continenti (si veda l'articolo *La conferma della deriva dei continenti* di Patrick M. Hurley, in « Le Scienze », n. 3, novembre 1968). Varie prove, derivanti dalla forma, dalle strutture e dalla paleontologia dei diversi continenti, nonché, negli ultimi vent'anni, i rilievi sul « paleomagnetismo » fissato nelle rocce eruttive, suggerirono che i continenti avessero raggiunto le loro attuali posizioni provenendo da posizioni relative molto differenti, nel corso di milioni di anni. Ma, benché fosse concepibile la possibilità di tali spostamenti, non era ancora molto chiaro quali forze potessero aver determinato movimenti così cospicui delle masse continentali.

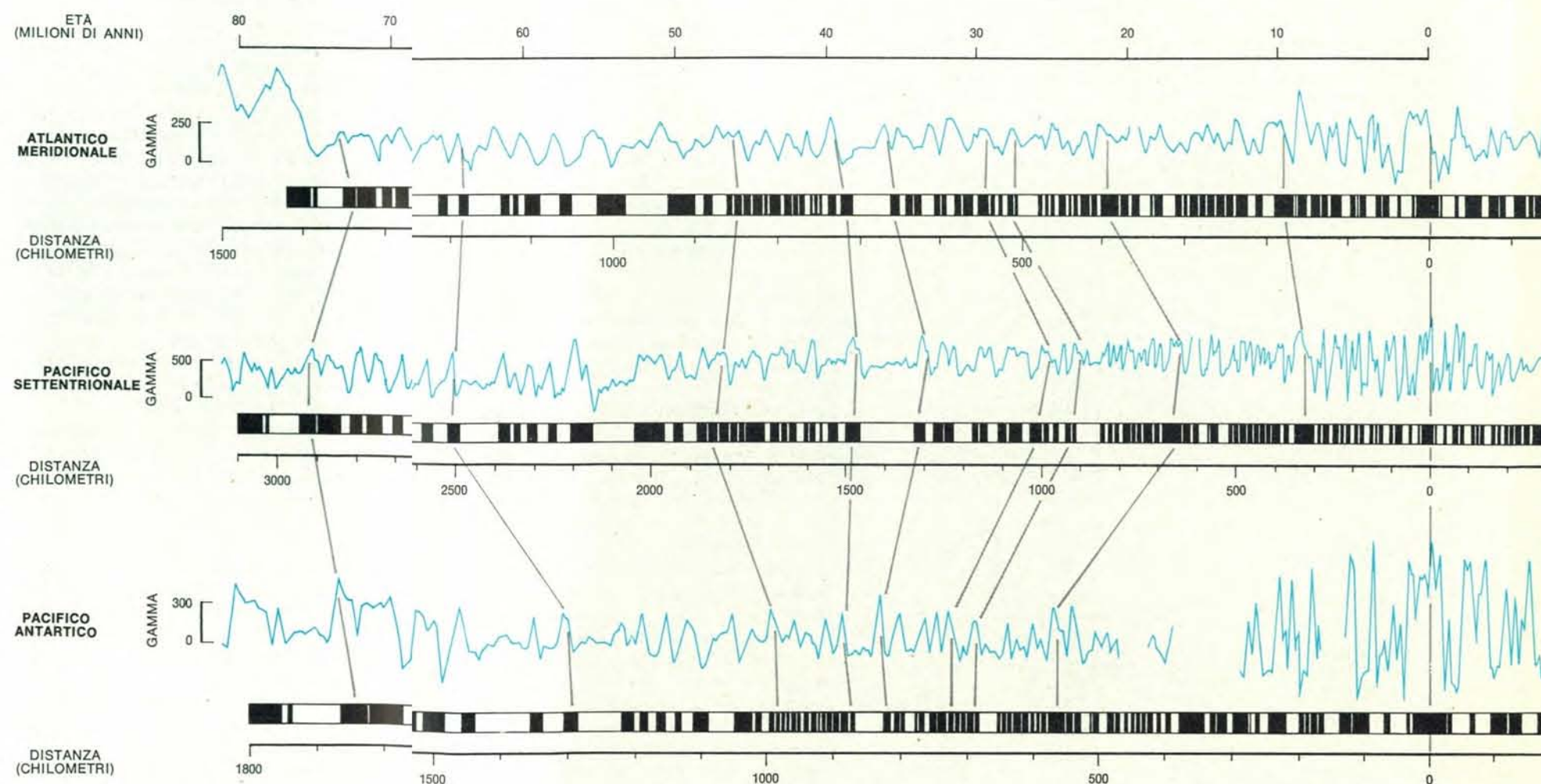
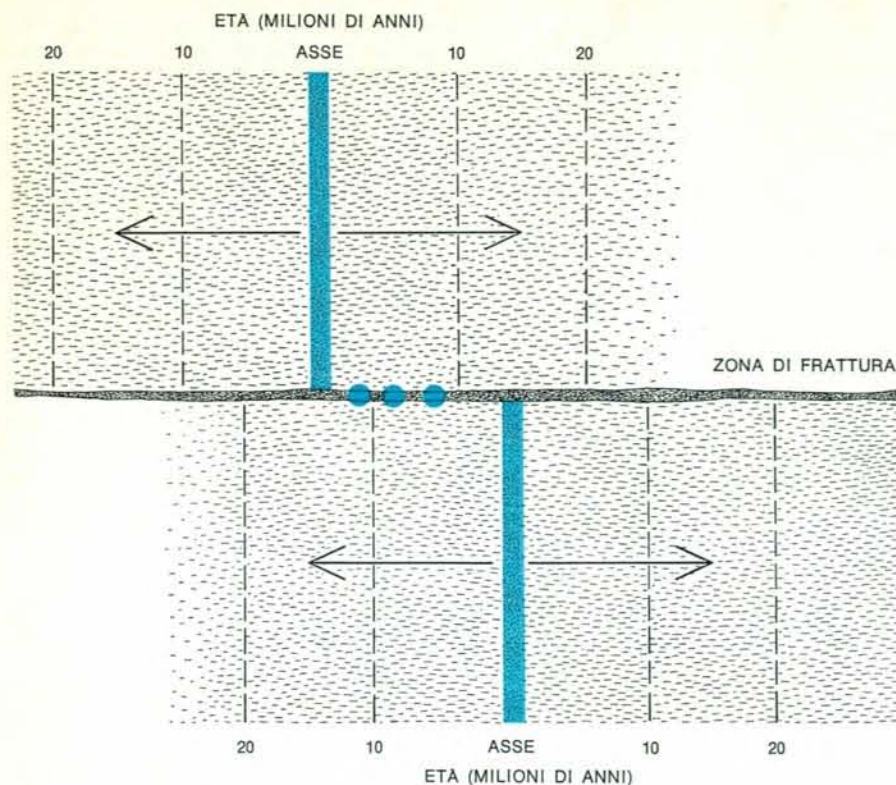
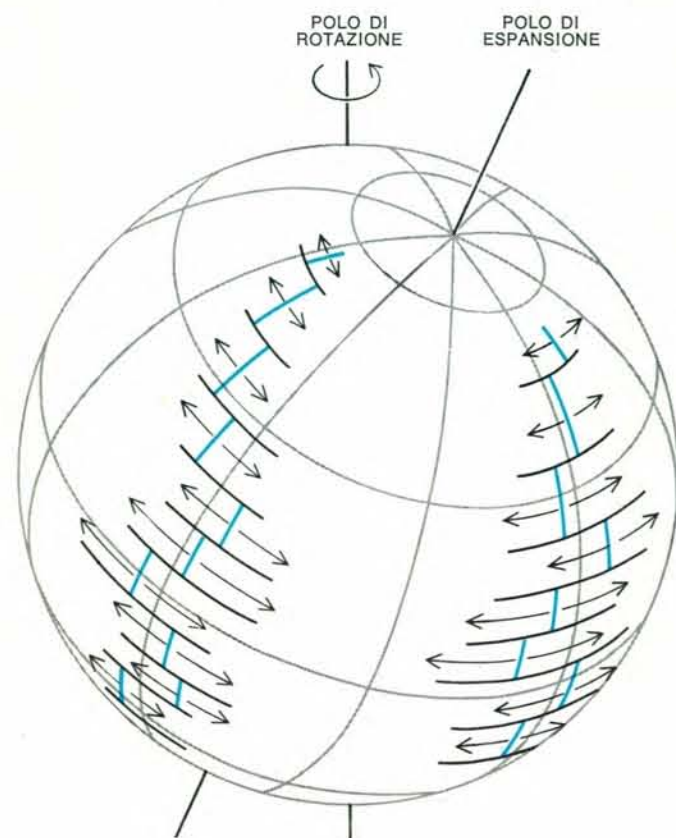
Sul finire degli anni cinquanta gli oceanografi accertarono che una dorsale sottomarina continua e sinuosa si estende e si dirama per tutti gli oceani, che questa dorsale si trova di solito lungo l'asse dei bacini oceanici e che a essa sono associate aree di intensa attività sismica. Dal canto loro i geologi constatarono che le rocce dei fondi oceanici erano molto recenti: nessun campione prelevato dal fondo marino risultò infatti anteriore al Cretacico, che ebbe inizio circa 135 milioni di anni fa. Verso il 1960 Harry H. Hess della Princeton University prospettò che i fondi oceanici potessero essere in movimento. Egli suggerì che un movimento di tipo convettivo potesse spingere il materiale profondo verso la superficie lungo l'asse delle dorsali oceaniche, quindi lo facesse espandere all'esterno attraverso il fondo dell'oceano e sprofondare in corrispondenza delle fosse oceaniche, ai margini delle masse continentali. Questa ipotesi appariva molto affascinante nel caso dell'Oceano Pacifico, che è contornato da fosse, ma risultava meno soddisfacente per gli altri oceani, che ne sono privi.

Pressoché nello stesso periodo, Ronald G. Mason, Arthur D. Raff e Victor Vacquier, della Scripps Institution of Oceanography, scoprirono che il fondo oceanico prospiciente la costa occidentale dell'America del nord mostrava variazioni dell'intensità magnetica disposte secondo fasce assai regolari. Tale schema indicava corpi magnetici molto estesi e sottili, orientati in direzione nord-sud, a tratti dislocati secondo allineamenti ortogonali. Non era mai stata osservata alcuna struttura che potesse spiegare uno schema di questo tipo; l'origine di questi singolari corpi magnetici rimase un mistero per quasi cinque anni. Nel 1963 F. J. Vine e D. H. Matthews, dell'Università di Cambridge, proposero una brillante verifica dell'ipotesi di Hess sui movimenti dei fondi oceanici basata sulla scoperta, confermata nei dettagli, che il campo magnetico terrestre aveva invertito la propria polarità diverse volte nei tempi geologici. Essi pensarono che le rocce fuse, risalite lungo l'asse delle dorsali oceaniche, si dovevano essere magnetizzate secondo la direzione del campo magnetico quale era al tempo della loro solidificazione: il materiale raffreddato, allontanandosi dalla dorsale, avrebbe così formato fasce alternate di magnetismo « normale », o attuale, e « invertito ». Un rilievo magnetometrico sulla superficie dell'oceano avrebbe dovuto quindi rilevare queste fasce come anomalie magnetiche, positive o negative, rispetto al campo magnetico terrestre esistente al livello del mare.

Conferma sperimentale

Mentre Vine e Matthews pubblicavano il loro lavoro, ero impegnato, assieme a colleghi del Lamont Geological Observatory della Columbia University e del U.S. Naval Oceanographic Office, nel rilevamento magnetico di dettaglio della dorsale di Reykjanes, facente parte della dorsale atlantica a sud dell'Islanda, nota per le sue notevoli anomalie magnetiche. Scoprimmo che le anomalie avevano andamento lineare e simmetrico, disposto parallelamente all'asse della dorsale. Ciò rappresentava una significativa conferma dell'ipotesi dell'espansione del fondo oceanico, con relativa formazione di anomalie magnetiche, proprio com'era stato suggerito da Vine e Matthews. Un poco più tardi lo stesso Vine e J. Tuzo Wilson dell'Università di Toronto mettevano in evidenza che le recenti inversioni del campo magnetico combaciavano, una per una, con parte degli estesi allineamenti magnetici registrati in prossimità della costa occidentale dell'America settentrionale da Mason, Raff e Vacquier.

Nel 1965 fu chiaro, a noi e ad altri, che il magnetismo poteva rappresentare la chiave per costruire la storia dei fondi oceanici e dei movimenti dei continenti. In soli tre anni si erano fatti notevoli progressi. Gli studiosi che hanno dato un contributo significativo a queste ricerche sono assai numerosi, al punto che è impossibile citarli tutti in un breve articolo, o anche semplicemente indicare chi furono i primi a fare nuove osservazioni o a proporre nuovi modelli.

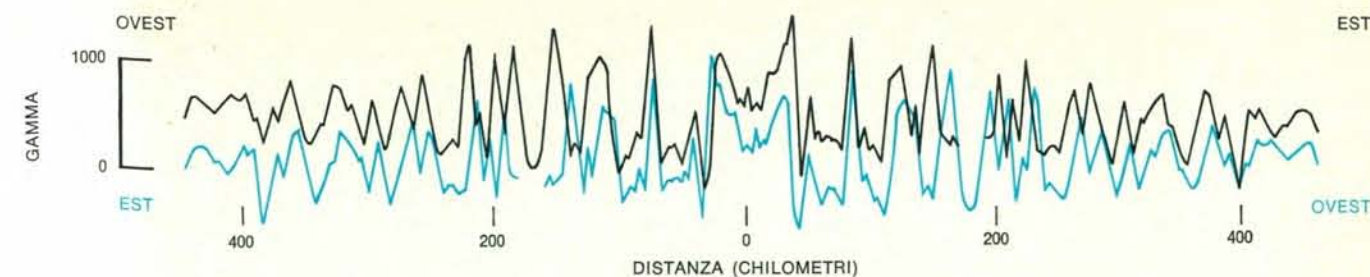
I primi studi sulle misure del campo magnetico sul mare furono iniziati al Lamont circa vent'anni fa. Strumenti semplici e precisi furono approntati per essere trainati dalle navi, mentre al tempo stesso si elaborarono tecniche efficienti per la registrazione e l'interpretazione dei dati necessari alla rico-

Le anomalie magnetiche (*linee colorate*) registrate in tutti gli oceani rivelano la medesima successione di corpi magnetizzati (*bande bianche e nere*), paralleli alle dorsali oceaniche. I corpi sono costituiti da rocce che hanno raggiunto la superficie in corrispondenza degli « assi di espansione », in periodi successivi caratterizzati da un campo magnetico « normale », come l'attuale, o « invertito ». I minerali componenti la roccia, inizialmente allo stato fuso, vengono magnetizzati nel senso del campo magnetico, quindi la roccia viene allontanata dall'asse di espansione da nuove risalite di materiale. Nella figura sono riportate le rilevazioni magnetiche relative a tre oceani. Le anomalie (espresse in gamma, che è una unità di misura dell'intensità del campo magnetico) e i corpi magnetici che a esse si riferiscono sono distanziati in modo diverso nei vari oceani per effetto delle differenti velocità di espansione, ma si ha sempre la medesima successione di 171 inversioni in 76 milioni di anni.

La roccia fusa risale dal profondo e raggiunge la superficie in corrispondenza all'asse di espansione, si solidifica e viene allontanata da nuove risalite di materiale. L'asse è dislocato da una zona di frattura. Nel tratto compreso fra le due porzioni dell'asse dislocato le masse rocciose si muovono, sui due lati della zona di frattura, in senso opposto, generando terremoti con ipocentro a piccola profondità (*dischetti colorati*).

La relazione geometrica fra l'asse delle dorsali e le zone di frattura risulta evidente se si immaginano le linee di latitudine e longitudine corrispondenti a un « polo di espansione ». In ogni oceano le zone di frattura risultano perpendicolari agli assi di espansione, mentre l'entità dell'espansione dei fondi oceanici (*frecce*) risulta massima all'equatore del nuovo sistema di coordinate, e diminuisce regolarmente verso i poli.

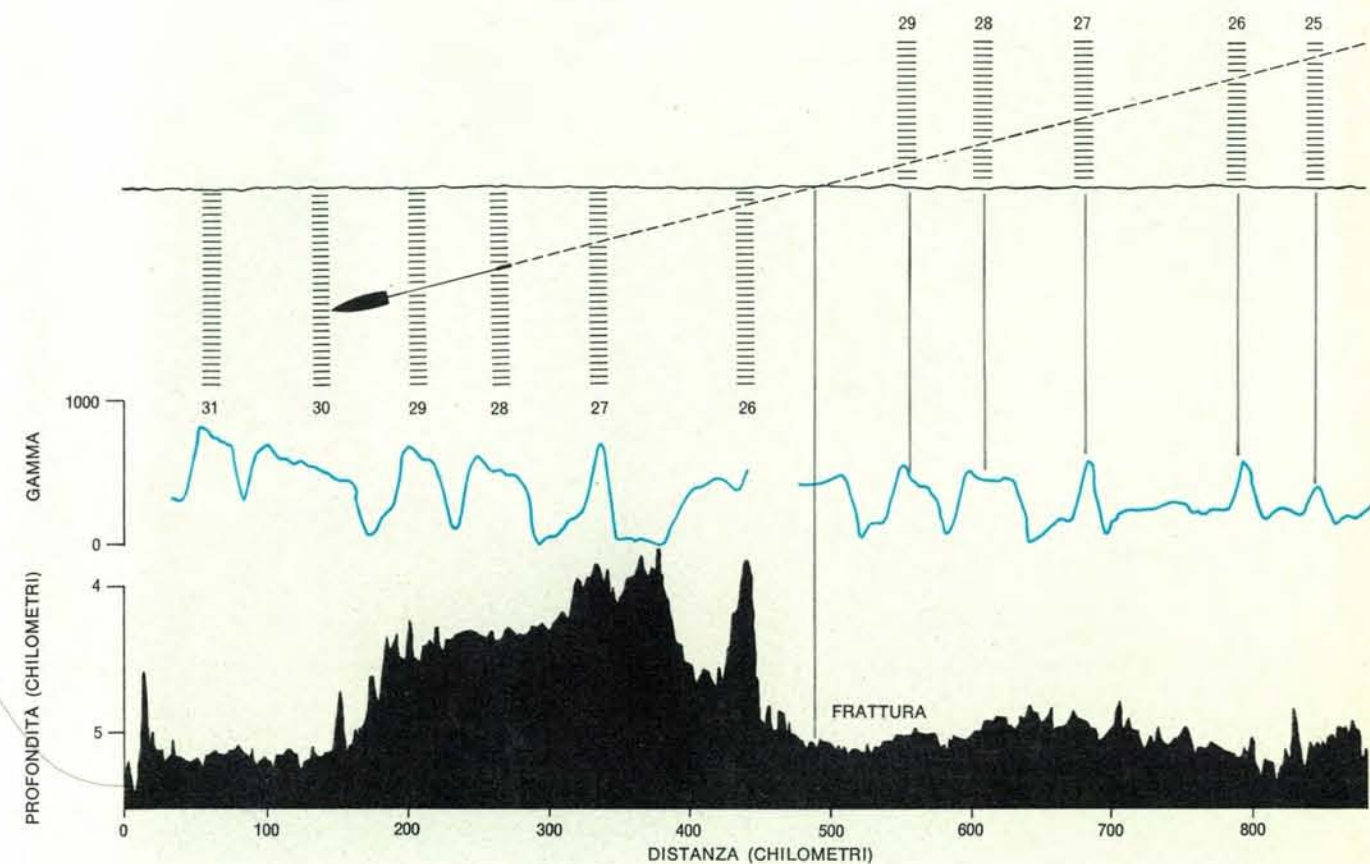
struzione delle strutture dei fondi oceanici sottostanti agli strati sedimentari. Nel 1965, quando si delineò compiutamente l'importanza delle anomalie magnetiche, disponevamo di una gran messe di dati rilevati in tutti gli oceani e anche delle tecniche di elaborazione dei dati con i calcolatori. Esaminando i dati sulla base delle nuove ipotesi, fummo in grado di riconoscere che la medesima successione di corpi magnetici si riscontrava in corrispondenza delle dorsali del Pacifico meridionale, dell'Atlantico meridionale e dell'Oceano Indiano.

Un ulteriore esame ha rilevato uno schema dell'espansione dei fondi oceanici che, su scala mondiale, conferisce un significato a un'ampia varietà di osservazioni. Esso, per esempio, sembra spiegare molti terremoti, stabilisce una scala cronologica dettagliata per le inversioni del campo magnetico, e spiega la direzione e l'entità della deriva dei continenti. I fenomeni geologici che intervengono nell'espansione lungo le dorsali e nello sprofondamento lungo i bordi dei continenti non sono tuttavia ancora compiutamente spiegati.

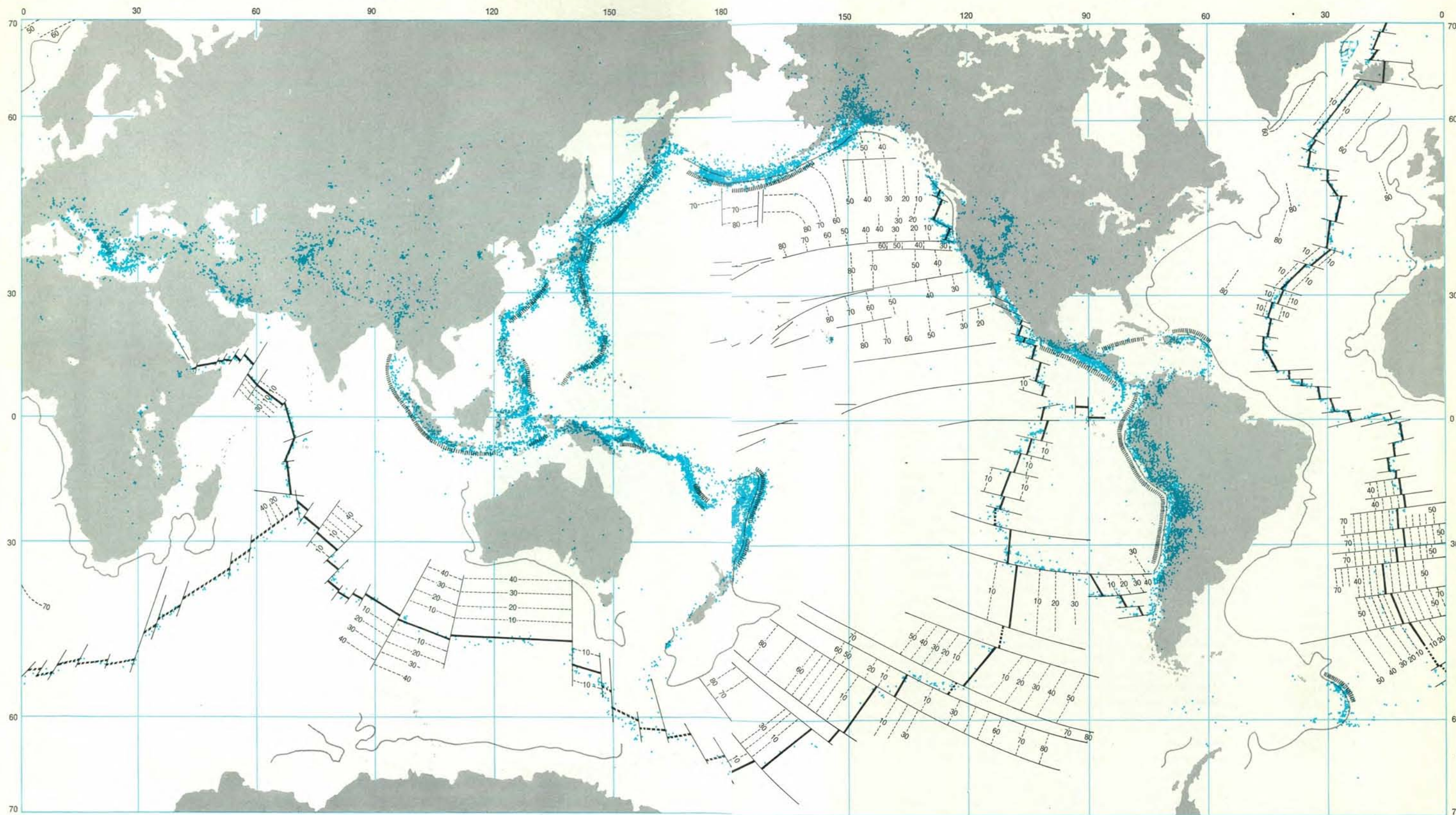
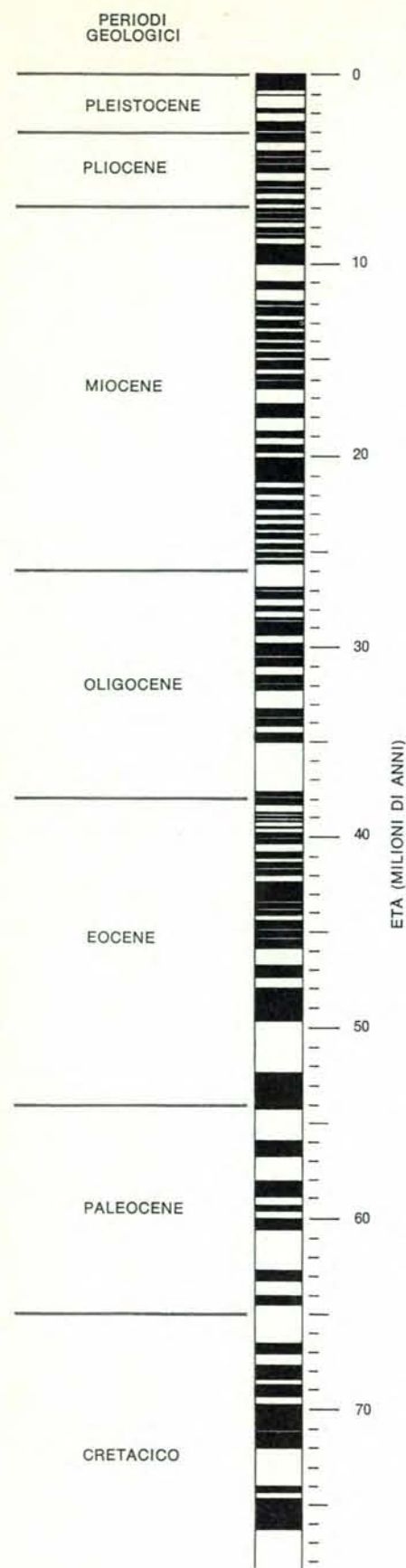
L'espansione su scala mondiale

A questo punto conosciamo i lineamenti principali dell'espansione in circa la metà delle aree oceaniche della Terra, per lo piú nelle zone adiacenti alle dorsali. Le aree per le quali lo schema di espansione dei fondi oceanici è ignoto, o sono inesplorate o si presentano di difficile interpretazione. Molti oceani, per esempio, non sono abbastanza vasti da mostrare i corpi magnetici piú antichi e non è possibile registrare un numero di anomalie magnetiche sufficiente a stabilire la velocità dell'espansione. Alcune aree, inoltre, non presentano anomalie magnetiche, il che può essere dovuto alla stabilità dei fondi oceanici, oppure al fatto che gli stessi risalgono a età geologiche molto antiche, durante le quali il campo magnetico terrestre potrebbe non aver presentato inversioni. Vi possono essere inoltre aree ove l'espansione è cosí lenta che i corpi magnetici, addossati uno all'altro, non permettono la risoluzione delle anomalie magnetiche. Abbiamo anche trovato aree in cui l'asse di espansione sembra essere al disotto dei continenti, o comunque non coincidente con l'asse delle dorsali oceaniche.

La dorsale oceanica, che costituisce l'asse dell'espansione, ha l'andamento di una linea spezzata essendo dislocata in molti punti. Le dislocazioni, o zone di frattura, presentano un'estensione considerevole su entrambi i lati della zona assiale e sono spesso indicate da alcune irregolarità nella topografia del fondale oceanico. Nella zona di frattura compresa fra i due tratti dell'asse dislocato, i movimenti delle masse rocciose avvengono in senso opposto (*si veda la figura in alto nella pagina a fronte*), con sviluppo di fenomeni che possono generare terremoti. I sismologi possono oggi localizzare l'ipocentro dei terremoti con l'approssimazione di pochi chilometri, come pure sono in grado di determinare la direzione del movimento iniziale fra le zolle della crosta terrestre che hanno prodotto il sisma. Le registrazioni sismiche confermano in pieno la teoria della espansione dei fondi oceanici, dal momento che la maggior parte dei terremoti con ipocentro in corrispondenza delle dorsali sono originati nella zona di frattura compresa fra i due tratti dell'asse dislocato, mentre la direzione del movimento iniziale corrisponde a quella prevista sulla base della teoria della espansione.

Anche se questo non si può vedere su una carta tracciata con la proiezione di Mercatore, le zone di frattura sono perpendicolari agli assi di espansione. Ciò risulta particolarmente evidente nel caso del Pacifico e dell'Atlantico meridionale. Se si tracciano le linee corrispondenti alla latitudine e alla longitudine rispetto a un « polo di espansione » e non al polo di rotazione, risulta che l'asse di espansione è parallelo alle nuove linee di longitudine tracciate, mentre le zone di frattura sono perpendicolari all'asse e parallele alle linee di latitudine (*si veda la figura in basso nella pagina a fronte*). Inoltre, l'espansione è maggiore in corrispondenza dell'equatore del sistema di coordinate cosí stabilito, mentre decresce verso i poli. Ne deriva pertanto che i fondi oceanici sono interessati da linee

La simmetria della registrazione magnetica deriva dal fatto che le rocce magnetizzate si sono allontanate su entrambi i lati dell'asse. La simmetria è dimostrata chiaramente dal confronto della registrazione (*in nero*) con la sua immagine speculare (*in colore*).

La zona di frattura può essere identificata dal confronto del profilo magnetico con quello del fondale oceanico. Il magnetometro, al traino di una nave, ha intersecato per due volte (*linea tratteggiata*) gli stessi corpi magnetici (*numeri 26-29*) dislocati da una zona di frattura. I corpi magnetici determinano anomalie dello stesso tipo, riconoscibili dal profilo magnetico (*al centro*). Il profilo batimetrico (*in basso*) indica che la zona di frattura è prossima a particolarità morfologiche significative.

di frattura di proporzioni gigantesche in corrispondenza delle dorsali. I poli di espansione sono diversi per i vari oceani. Per il Pacifico e l'Atlantico meridionale essi sembrano assai prossimi ai poli magnetici, trovandosi nei pressi della Groenlandia e nell'Antartide a sud dell'Australia. L'espansione dell'Oceano Indiano sembra riferirsi a un'altra coppia di poli che si possono ubicare, con minore precisione, nell'Africa del nord e nel Pacifico a nord della Nuova Zelanda. Le informazioni attualmente disponibili indicano che l'espansione dell'Atlantico settentrionale avverrebbe rispetto a una terza coppia di poli, mentre le zone di frattura delle parti piú antiche del Pacifico nordorientale sarebbero associate con un'ulteriore coppia di poli. Vi sarebbero inoltre indicazioni che alcuni, e forse tutti gli assi di espansione sono essi stessi in movimento e che pertanto lo stesso si può dire dei poli relativi.

Alcuni studiosi, e particolarmente W. Jason Morgan di Princeton, sono inclini a generalizzare i movimenti della crosta terrestre, ipotizzando che essa sia suddivisa in almeno sei placche rigide. Queste placche si accrescono per l'ag-

La rappresentazione su scala mondiale dell'espansione dei fondi oceanici risulta particolarmente significativa se si confrontano i dati magnetici con quelli sismici. Le dorsali oceaniche (*tratti neri marcati*) sono dislocate trasversalmente dalle zone di frattura (*linee sottili*). In base ai valori dell'espansione, stabiliti sui dati magnetici, l'autore e i suoi colleghi hanno determinato le linee «isocrone», che rappresentano l'età dei fondi oceanici in milioni di anni (*linee tratteggiate sottili*). Ai margini di molte masse continentali (*linee grige*) si trova buona parte delle fosse oceaniche profonde (*tratteggiate*). Riportando gli epicentri dei terremoti registrati tra il 1957 e il 1967 (*punti colorati*) essi si addensano presso le dorsali o le fosse oceaniche.

La cronologia delle inversioni del campo magnetico terrestre è stata ricavata dall'estrapolazione dei dati relativi alle inversioni piú recenti, assumendo valori di espansione costanti per ogni oceano. I risultati sono confortati dai dati geologici.

giunta di nuovo materiale, in corrispondenza degli assi di espansione, mentre i margini esterni delle placche si sormontano l'uno con l'altro. Il concetto delle placche rigide risulta attraente per i sismologi poiché, come verrà spiegato in seguito, i terremoti profondi risultano provenire da zone che corrispondono ai margini di molte di esse.

La cronologia geomagnetica

I corpi magnetizzati dei fondi oceanici forniscono la storia, sorprendentemente completa, delle inversioni del campo magnetico per un arco di tempo di 76 milioni di anni, fino a comprendere parte del Cretacico. Si iniziò con la scala cronologica stabilita dagli studiosi della Stanford University e dell'U.S. Geological Survey, che correlarono la magnetizzazione con l'età delle rocce, determinata con metodi radioattivi, risalendo a circa 3,5 milioni di anni. Confrontando le età assegnate a specifiche inversioni del campo magnetico con le distanze dall'asse delle dorsali, fummo in grado di estendere le osservazioni di Vine e Wilson a buona parte dei fondi oceanici, determinando le velocità di espansione nei vari oceani. Le velocità risultarono diverse per vari tratti della dorsale, ma in molte aree sembravano assai costanti nel tempo. Esse variano da due a cinque centimetri all'anno. Si tratta di valori notevoli su scala geologica, anche se risultano dello stesso ordine di quelli riscontrati lungo la faglia di S. Andrea nella California meridionale.

Nelle diverse aree studiate non abbiamo rilevato alcuna evidente interruzione nel movimento di espansione. Abbiamo pertanto ritenuto di poter assumere come costante la velocità di espansione e di datare, su questa base, le inversioni della polarità magnetica ben oltre i 3,5 milioni di anni della scala cronologica citata. Questa estrapolazione può apparire a prima vista ingiustificata, ma essa è in effetti suffragata dall'accordo dei dati ottenuti nei vari oceani, nonché dalla compatibilità con altre prove geofisiche, quali per esempio le datazioni radioattive e paleontologiche delle rocce. Nei limiti di errore insiti nei vari metodi, non si sono rilevate discordanze. Finora abbiamo identificato 171 inversioni del campo magnetico terrestre su un arco di tempo di 76 milioni di anni. Riteniamo che le datazioni di ciascuna di queste fasi magnetiche siano abbastanza accurate. Evidentemente, se l'espansione dei fondi oceanici si fosse interrotta bruscamente in tutto il mondo per un certo periodo di tempo, la nostra scala cronologica risulterebbe accorciata dello stesso periodo di tempo. Tale eventualità non sembra tuttavia molto probabile, anche se non può essere esclusa del tutto. Nel complesso, le prove dell'accuratezza della scala cronologica sono talmente rilevanti che essa può essere ora utilizzata per lo studio delle variazioni delle velocità di espansione in alcune delle aree che risultano piú complesse.

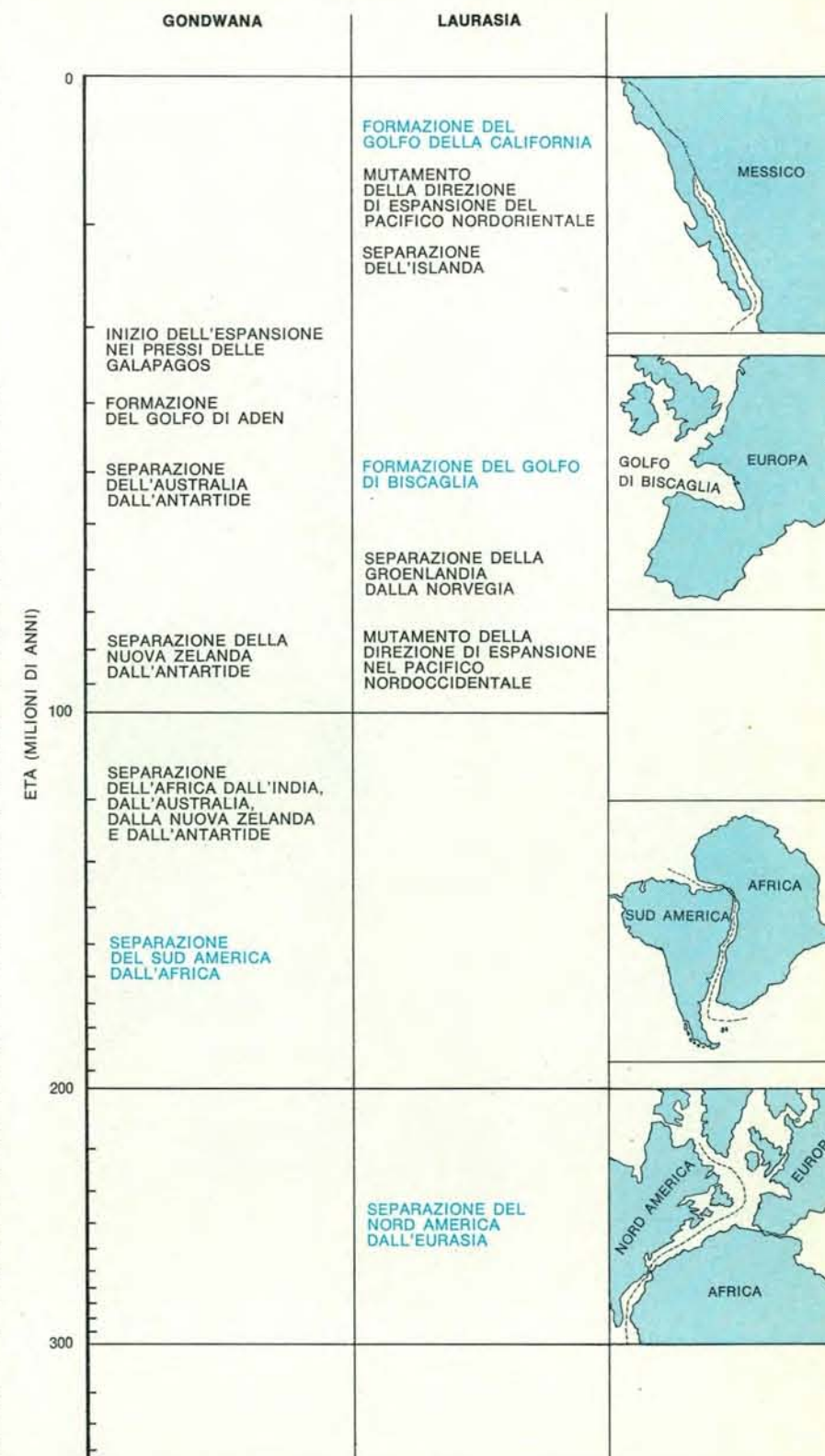
La durata media dei periodi con campo magnetico normale, o attuale, risulta di 420 000 anni, mentre quella dei periodi con campo magnetico invertito è di 480 000. I due valori, abbastanza vicini, indicano come la Terra abbia all'incirca la stessa probabilità di trovarsi in uno stato magnetico e nell'altro. Il periodo di polarità magnetica attuale è già durato 700 000 anni e pertanto siamo forse prossimi a un cambiamento. Soltanto il 15 % dei periodi a polarità normale sono infatti durati piú di 700 000 anni, anche se pare che alcuni periodi abbiano raggiunto i tre milioni di anni. D'altro canto, i periodi piú brevi risulterebbero inferiori ai 50 000 anni, ma in realtà è difficile controllare questo dato, in quanto si tratta di intervalli di tempo troppo brevi per essere confermati mediante metodi di datazione assoluti. Ciò lascia intravedere uno dei limiti della scala cronologica geomagnetica, che per essere cosí dettagliata difficilmente potrà essere presto smentita, mentre d'altro canto è difficile usarla per datare corpi magnetici limitati. Per utilizzare la scala cronologica geomagnetica è necessario conoscere l'età approssimativa dei materiali, cosí come si usa il mirino telescopico per orientare un potente telescopio.

Il geomagnetismo di cui abbiamo parlato è «congelato» nelle rocce ignee di tipo basaltico che risalgono dal profondo della Terra. Naturalmente, su gran parte del fondo oceanico questo basamento igneo è ricoperto da depositi sedimentari di vario spessore. Anche i sedimenti possono essere magnetizzati nel senso del campo magnetico terrestre quando si depositano sul fondo dell'oceano. (La loro magnetizzazione rappresenta tuttavia soltanto un decimillesimo di quella dei basalti, cosicché anche un notevole spessore di sedimenti non interferisce con le misure delle anomalie magnetiche determinate dalle rocce del basamento.) Carotando il fango dei fondi marini, si possono portare alla superficie campioni degli strati sedimentari che registrano le inversioni del campo magnetico. Alcuni studiosi hanno recentemente sviluppato delle tecniche assai sensibili per misurare la debole magnetizzazione di questi campioni. In tal modo sono state registrate le inversioni magnetiche fino a circa dieci milioni di anni fa. La possibilità di correlare nel tempo strati sedimentari dei diversi oceani si è dimostrata di grande utilità per i geologi, che spesso non dispongono di altre valide indicazioni sull'età degli strati. Si sono osservate evidenti correlazioni fra le inversioni magnetiche e le principali modificazioni delle microfaune fossili. Si è pensato che tali modificazioni fossero il risultato di mutazioni determinate dall'aumento dell'esposizione ai raggi cosmici, qualora l'effetto protettivo del campo magnetico terrestre si fosse attenuato nel corso dell'inversione di polarità magnetica. Tuttavia vi possono essere spiegazioni alternative, che verranno esposte successivamente in questo articolo.

Tracce lasciate dai continenti

I movimenti dei fondi oceanici sono compatibili, come dimensioni e direzioni, con la deriva dei continenti. Le prove geologiche e topografiche hanno indicato che, due o trecento milioni di anni fa, esisteva una sola grande massa continentale nell'emisfero meridionale (denominata Gondwana da una tipica provincia geologica dell'India) che comprendeva l'Africa, l'India, l'America meridionale, l'Australia, la Nuova Zelanda e l'Antartide, e un'altra massa continentale nell'emisfero settentrionale, denominata Laurasia. Le posizioni dei continenti attuali nell'ambito di queste grandi masse continentali non erano tuttavia molto chiare e nemmeno era possibile tracciare in dettaglio la sequenza degli eventi succeduti alla disgregazione delle due grandi masse continentali.

Gli allineamenti magnetici dei fondi oceanici costituiscono le tracce lasciate dai continenti, segnando le loro successive posizioni occupate prima di raggiungere quelle attuali. Abbiamo riscontrato che il lento, ma continuo e prolungato, spostamento dei fondi oceanici fu sufficiente a distanziare l'America meridionale dall'Africa in circa duecento milioni di anni, creando per tutta la sua estensione l'Atlantico meridionale, e a separare l'Australia dall'Antartide in circa quaranta milioni di anni. Con la datazione di altri fondi oceanici, si poté stabilire piú esattamente quando i vari continenti si separarono e come avvennero gli spostamenti (*si veda l'illustrazione a destra*). È ora possibile ricostruire le posizioni originali dei continenti e delle piattaforme continentali, cosicché i geologi possono controllare la continuità delle strutture geologiche che furono separate dall'allontanamento delle masse continentali.

Sebbene sia possibile dire come e quando i continenti si separarono uno dall'altro, non è sempre del tutto chiaro se vi siano state interruzioni nel movimento di allontanamento. Si ha l'impressione che sia l'Africa sia l'Antartide abbiano mantenuto la loro posizione rispetto all'asse di rotazione negli ultimi 100 milioni di anni, o comunque da quando l'Africa si separò dagli altri frammenti del Gondwana. Se ciò è vero, il fatto che l'asse di espansione fra l'America meridionale e l'Africa si mantenga in posizione intermedia fra i due continenti indica che anche l'asse si deve essere spostato.

La scala cronologica delle modificazioni della crosta terrestre si può ora determinare in base alle isocrone, che indicano l'età dei fondi oceanici e le direzioni di espansione.

Origine e sprofondamento dei materiali

Fino alla formulazione della teoria sull'espansione dei fondi oceanici, le sole fonti note di materiali provenienti dal profondo della Terra erano costituite dalle eruzioni vulcaniche. I vulcani piú attivi nell'epoca attuale non sono tuttavia ubicati in corrispondenza delle zone di sollevamento del materiale subcrostale, corrispondenti alle dorsali oceaniche, ma piuttosto nelle aree ove i fondi oceanici in movimento si spingono sotto le masse continentali attuali. Le recenti eruzioni vulcaniche delle Filippine, del Messico e del Guatemala, come pure le continue attività vulcaniche del Nord America occidentale, dell'Alaska e del Giappone, rappresentano al riguardo degli esempi tipici. Il sollevamento che dà origine all'espansione dei fondi oceanici rappresenta pertanto un fenomeno geofisico finora praticamente sconosciuto, che ha concentrato l'attenzione dei ricercatori sull'asse delle dorsali oceaniche, ove esso si manifesta. L'asse presenta diverse caratteristiche non comuni: un grande flusso di calore, un'elevata concentrazione di terremoti poco profondi, velocità sismiche non usuali, mancanza di coperture sedimentarie, una rilevante anomalia magnetica. Mentre la maggior parte di queste caratteristiche non comuni si estende per decine o centinaia di chilometri su entrambi i lati delle dorsali, l'anomalia magnetica è invece nettamente localizzata, al punto che le rilevazioni magnetiche permettono di ubicare l'asse delle dorsali con l'approssimazione di pochi chilometri. La topografia delle dorsali è molto accidentata e non è possibile associare l'asse di espansione con alcuna delle singole particolarità topografiche. Si può dire tuttavia che, nei casi in cui la dorsale comprende una fossa mediana, l'asse risulta immancabilmente in corrispondenza di questa.

Le antiche posizioni dei continenti possono essere ricostruite sulla base della scala cronologica. Questa carta mostra le posizioni relative del continente americano con l'Eurasia e l'Africa, quali erano 70 milioni di anni fa. Si pensa che i continenti combaciassero al limite delle masse continentali (*grigio chiaro*), piuttosto che ai limiti delle terre emerse attuali (*grigio scuro*). Le linee tratteggiate indicano la direzione del movimento.

La piastra di materiale magnetizzato non è molto spessa. L'analisi delle anomalie magnetiche suggerisce spessori da uno a pochi chilometri, mentre le indagini sulla trasmissione delle onde sismiche e sulla distribuzione del flusso di calore indicano che lo spessore entro il quale avvengono i movimenti convettivi risulterebbe di alcune decine di chilometri, con un massimo di circa 150. La linearità dei corpi magnetici indica che il materiale non ha subito ribaltamenti dopo il raffreddamento, a differenza di quanto accade nelle aree continentali corrugate, o nelle zone vulcaniche caratterizzate dalla irregolare sovrapposizione di colate di lava. Le rilevazioni effettuate immergendo un magnetometro fino nei pressi dei fondi oceanici hanno segnalato l'esistenza di corpi molto magnetizzati in corrispondenza dell'asse di espansione. Ciò suggerisce che l'emersione delle masse profonde si manifesti con l'iniezione di materiale fuso in corpi ad andamento lineare, denominati filoni. Tali corpi hanno un'elevata probabilità di essere iniettati nella fascia assiale delle dorsali (che rappresenta una zona di minor resistenza della crosta terrestre per effetto della spinta ascensionale della corrente convettiva). Il nuovo materiale ha una temperatura sufficientemente elevata da fondere la roccia incassante, cosicché il materiale di nuova iniezione e quello circostante vengono magnetizzati secondo la direzione del campo magnetico. Questa spiegazione non fornisce indicazioni precise sullo spessore dello strato entro il quale avvengono i movimenti, ma tuttavia indica come le masse rocciose magnetizzate non subiscano ribaltamenti. I filoni tendono infatti a spingere il materiale incassante in senso orizzontale.

Le profonde fosse che si trovano ai margini dell'Oceano Pacifico sono ritenute le aree di sprofondamento del materiale del fondo oceanico. Lo stesso avviene probabilmente nelle fosse dello Atlantico e dell'Oceano Indiano. In molte parti del globo, dove sono state identificate le placche in movimento del fondo oceanico, non si sono potuti localizzare i margini esterni, cosicché non è possibile riconoscere l'ubicazione di tutte le aree di sprofondamento. In molte zone il fondo oceanico sembra ripiegarsi al disotto delle masse continentali, ma in alcuni luoghi sembra respingere i continenti, ovvero separarli l'uno dall'altro. Non sono note invece aree nelle quali il fondo oceanico in espansione sormonta le masse continentali.

Se si riportano gli ipocentri dei terremoti profondi, si nota che essi si trovano quasi tutti su un piano che, disposto secondo l'allineamento del fondo di una fossa, s'immerge con un angolo di circa 45° (*si veda la figura in alto a pagina 78*). La causa dei terremoti è rappresentata dallo scorrimento del materiale, che si estende per centinaia di chilometri al disotto della superficie, parallelamente al piano stesso. I piani che comprendono l'ipocentro dei terremoti delimitano quasi certamente le aree di sprofondamento. Lo studio del movimento iniziale delle masse che hanno generato il sisma indica che in queste aree il fondo oceanico presenta un movimento discendente rispetto alle masse continentali adiacenti. Al disotto dei sedimenti oceanici si è riscontrato uno strato relativamente sottile, con caratteristica velocità di propagazione delle onde sismiche, che scende a grandi profondità con una pendenza di 45°.

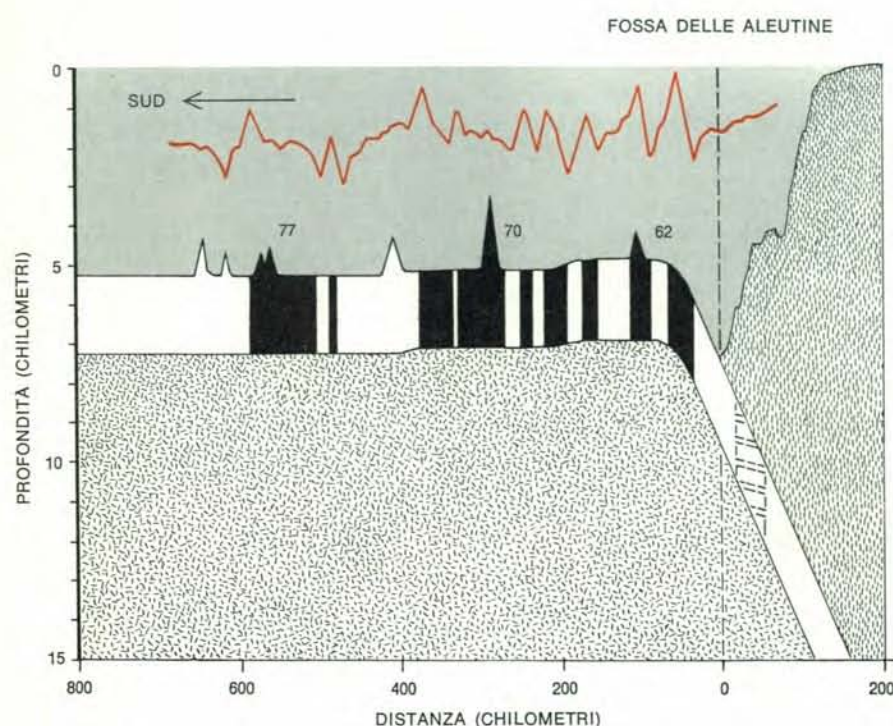
Non è invece chiaro, dal punto di vista dei rilievi magnetici, ciò che succede in corrispondenza delle fosse oceaniche. Le rilevazioni effettuate nella fossa delle Aleutine, per esempio, mostrano delle evidenti incompatibilità con gli schemi fin qui riportati (*si veda l'illustrazione in basso a pagina 78*). Non vi è infatti alcuna traccia di un corpo magnetico che, sulla base delle misure effettuate altrove, si dovrebbe trovare a una profondità di circa tre chilometri dal fondo della fossa. La scomparsa del corpo magnetico si potrebbe spiegare con la rifusione del materiale o con la smagnetizzazione dovuta a deformazioni meccaniche, ma di entrambi questi fenomeni non si rileva alcuna traccia a cosí piccola profondità dal fondo della fossa. Un altro problema che si riscontra nella fossa delle Aleutine è che i corpi magnetici sembrano assumere una successione inversa. Se infatti le fosse sono aree di sprofondamento, i corpi piú antichi si dovrebbero trovare in corrispondenza di queste e i piú recenti verso il mare aperto. Senonché si verifica esattamente l'opposto. A sud dell'Alaska, inoltre, i corpi magnetici orientati all'incirca nord-sud si piegano e assumono un andamento est-ovest, parallelo alla costa. Queste caratteristiche sono difficili da spiegare, se si ammette che le fosse rappresentino aree di sprofondamento, e rappresentano uno degli ostacoli piú difficili da superare per l'accettazione della nuova teoria.

Considerazioni conclusive

La sensazione di molti geofisici che si stia per giungere a teorie globali sui fenomeni terrestri deriva dalla singolare coincidenza nello spazio e nel tempo di certi fenomeni geofisici, molti dei quali riguardano le dinamiche del nostro pianeta. Le teorie esistenti non stabiliscono infatti alcuna relazione di causa ed effetto tra questi fenomeni.

Riportiamo brevemente alcune di queste coincidenze. Il polo di espansione attuale si trova, almeno per diversi oceani, in prossimità dell'asse magnetico terrestre. Analogamente il polo di espansione del Pacifico settentrionale, nel Cretacico, si trovava in prossimità del polo magnetico dello stesso periodo. In corrispondenza delle inversioni magnetiche si sono spesso notati significativi cambiamenti nelle microfaune marine. La caduta di un grande meteorite avvenne esattamente in corrispondenza dell'ultima inversione del campo magnetico. Alcuni studiosi hanno recentemente preso in considerazione i rapporti esistenti fra l'attività orogenetica e le inversioni magnetiche, mentre altri intravedono una correlazione fra le variazioni nell'espansione dei fondi oceanici e le stesse fasi orogenetiche. È anche stata suggerita la possibilità che il campo magnetico terrestre sia generato da movimenti convettivi, causati a loro volta da irregolarità dell'orbita terrestre. È anche stata ripresa una teoria, vecchia di trent'anni, secondo la quale le epoche glaciali furono causate da cambiamenti dell'inclinazione dell'asse terrestre. Vi sono infine prove evidenti che i principali terremoti avvengano in corrispondenza di certe variazioni nel movimento di rotazione della Terra.

Questo articolo non è il luogo adatto per una completa valutazione di tutti questi aspetti. È tuttavia interessante

Il fondo oceanico in movimento si piega verso il basso nelle fosse che si trovano ai margini dell'Oceano Pacifico. Gli ipocentri dei terremoti profondi sono allineati al contatto fra la corrente di materiale discendente e la massa continentale. I punti colorati indicano gli ipocentri dei sismi registrati nella fossa di Tonga, nel Pacifico meridionale.

La fossa delle Aleutine fornisce dei dati che non si inquadrano facilmente nella teoria. La registrazione magnetica (*linea colorata*) indica la successione dei corpi a magnetizzazione normale e invertita (*bande nere e bianche*) in prossimità della fossa. Non vi è alcuna indicazione dell'esistenza del corpo magnetico che si dovrebbe trovare tre chilometri sotto il piano della fossa (*banda a piccoli tratti*). Inoltre i corpi magnetici risultano piú giovani, e non piú vecchi, in direzione della fossa (da 77 a 62 milioni di anni).

notare come un tratto comune di queste, e di altre considerazioni che si trovano al limite del campo di applicazione delle ricerche geofisiche, sia rappresentato dal ruolo tenuto dagli spostamenti dell'asse di rotazione terrestre. Sembra che variazioni anche minime possano influire in modo sorprendente sia sugli eventi climatici che si svolgono sulla superficie della Terra, sia sulle forze agenti all'interno di essa.

L'intima relazione fra il polo di espansione e quello magnetico suggerisce che i movimenti convettivi all'interno della Terra e il campo magnetico terrestre possano avere una causa comune. Entrambi potrebbero derivare sia da irregolarità orbitali sia da fenomeni di induzione nell'ambito della massa terrestre. Poiché il verso delle correnti convettive non muta quando il campo magnetico si inverte, è chiaro che i movimenti convettivi non generano per se stessi il campo magnetico, e nemmeno è probabile che l'inversione del campo possa determinare il movimento nelle cellule di conversione. Qualunque sia la forza motrice di questi due fenomeni, si direbbe che essa è legata ai movimenti della Terra. Si è recentemente dimostrato che i terremoti di intensità superiore a 7,5 nella scala di Richter, causano, o sono causati da cambiamenti della nutazione, un piccolo movimento rotatorio dell'asse di rotazione terrestre. Qualunque sia il meccanismo di questi cambiamenti, si può pensare che anche nel passato le variazioni dei movimenti assiali della Terra possano aver causato i principali terremoti, la formazione delle catene montuose e anche le inversioni del campo magnetico.

Concludendo, a intervalli di pochi mesi si verificano cambiamenti nel movimento di rotazione della Terra che interessano l'espansione dei fondi oceanici e causano i terremoti a essa associati. Se tali cambiamenti sono abbastanza ampi, si può anche arrivare a un'inversione del campo magnetico. Si noti che la presenza del campo magnetico e l'espansione dei fondi oceanici sembrano dovuti al semplice fatto che la Terra ruota su se stessa. Sono soltanto le variazioni di moto che sono associate a certi terremoti e alle inversioni del campo magnetico. I cambiamenti delle microfaune marine sono collegabili a variazioni climatiche, che a loro volta derivano da variazioni del moto terrestre.

Queste considerazioni non possono ancora essere confermate, ma nemmeno decisamente respinte. In verità, esse non sono piú astruse di quanto si presentassero, fino a pochi anni fa, le teorie sull'espansione dei fondi oceanici.

Laser ad anidride carbonica

È stato possibile ottenere fasci laser piú potenti e continui impiegando quale mezzo attivo molecole di anidride carbonica a bassa pressione.

di C.K.N. Patel

Sino a poco tempo fa si pensava che i piú potenti laser realizzabili fossero quelli a stato solido, per il semplice motivo che nei solidi le particelle responsabili del fenomeno sono molto piú addensate di quanto non accada nei gas. Ci si rese però subito conto che i laser a stato solido non erano privi di difetti. Essi offrivano infatti scarse prestazioni sia dal punto di vista della purezza spettrale sia da quello della coerenza spaziale del fascio di luce emesso. Inoltre, la maggior parte dei laser a stato solido ad alta potenza funziona solamente a impulsi; in altre parole, la loro emissione di potenza avviene sotto forma di lampi di luce corti e intensi e non come un fascio continuo. I primi laser a gas fornivano invece un fascio continuo con notevole purezza spettrale ed eccellente coerenza spaziale, anche se la potenza emessa era molto bassa rispetto a quella fornita dai laser a stato solido.

La situazione cambiò radicalmente con l'avvento dei laser a gas molecolari. Il piú rappresentativo fra i laser di questa nuova categoria è il laser ad anidride carbonica, che è in grado di generare un fascio continuo con una potenza di uscita di alcuni chilowatt e, nello stesso tempo, di mantenere l'alto grado di purezza spettrale e di coerenza spaziale caratteristico dei laser a gas atomico a potenza piú bassa. Recentemente è stato impiegato un laser ad anidride carbonica per ottenere un fascio infrarosso con una potenza di emissione di 8,8 chilowatt, che costituisce il piú potente fascio laser continuo finora ottenuto. L'importanza di una cosí alta emissione di potenza può essere efficacemente dimostrata dal fatto che un fascio infrarosso di alcuni chilowatt, focalizzato, è in grado di tagliare una lastra di acciaio di sei millimetri di spessore in pochi secondi (*si veda la figura qui a sinistra*).

Il fascio infrarosso generato dal laser ad anidride carbonica del Redstone Arsenal (*si veda la fotografia a pag. 82*) può perforare una lastra di acciaio inossidabile, spessa sei millimetri, in 10 secondi. Il fascio infrarosso, dopo essere passato attraverso la finestra Brewster, disposta obliquamente a un'estremità del tubo laser (*a sinistra*), e uno specchio semiriflettente, incide su uno specchio concavo (*a destra*) che lo focalizza sulla lastra d'acciaio (*al centro*). Le proprietà termiche dei laser ad anidride carbonica ad alta potenza, pur essendo molto promettenti dal punto di vista industriale, sono considerate secondarie per i possibili impieghi di questi laser nei sistemi di comunicazione ottica.

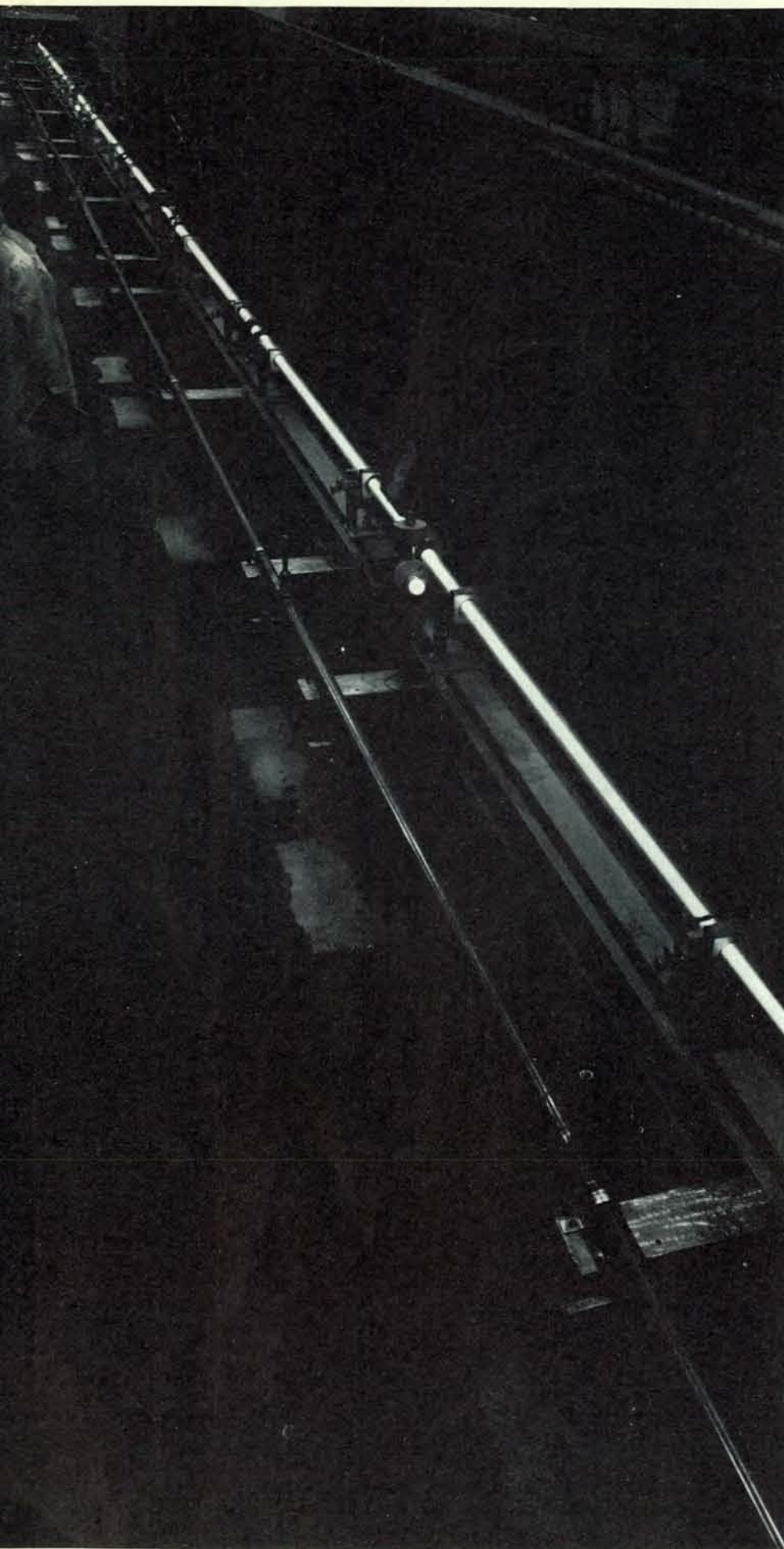
Con la loro alta potenza di emissione nella regione infrarossa dello spettro elettromagnetico, i laser a gas ad anidride carbonica hanno reso disponibile tutta una nuova gamma di lunghezza di onda per lo studio delle interazioni ottiche non distruttive con i gas, i liquidi e i solidi. Queste interazioni ottiche comprendono i processi non lineari, grazie ai quali è possibile generare una sorgente coerente di radiazioni infrarosse, regolabile con continuità su un ampio intervallo di frequenze. Esistono inoltre molte altre applicazioni per le quali i laser ad anidride carbonica ad alta potenza potranno essere molto utili. Fra queste applicazioni, quella che forse ha la maggior importanza potenziale va ricercata nel campo delle comunicazioni e dei radar ottici. Il laser ad anidride carbonica è particolarmente adatto per essere impiegato nei sistemi di comunicazioni sia terrestri sia extraterrestri, poiché il fascio infrarosso da esso prodotto è assorbito molto debolmente dall'atmosfera. In questo articolo cercherò di spiegare il principio fisico su cui si basa il funzionamento di questo nuovo tipo di laser a gas ad alta potenza.

Da un punto di vista generale, un laser a gas è costituito da un tubo (chiamato tubo laser) riempito di gas a bassa pressione e chiuso alle estremità da due specchi che formano una « cavità ottica ». Il gas (detto materiale laser) che riempie il tubo può essere formato da atomi, da vapori metallici o da molecole. L'effetto laser è normalmente ottenuto sottoponendo il gas a una scarica elettrica; gli elettroni liberati e accelerati durante la scarica urtano le particelle del gas, eccitandole a livelli energetici piú elevati, dai quali ritornano spontaneamente verso i livelli piú bassi. In seguito a questa diseccitazione, l'eccesso di energia viene emesso sotto forma di fotoni, o quanti di luce. Per ottenere il « guadagno » ottico caratteristico dell'effetto laser è necessario che la « densità di popolazione » delle particelle che si trovano nei livelli piú elevati sia maggiore di quella relativa ai livelli piú bassi. Questa condizione è conosciuta come « inversione di popolazione » perché è l'inverso della condizione normale delle particelle, che è la condizione di non eccitazione. Per ottenere un'elevata emissione di potenza in una transizione fra una coppia di livelli energetici, è necessario che il numero assoluto degli atomi eccitati al livello laser superiore sia grande e che le particelle gassose lascino il livello laser inferiore con la stessa velocità con la quale esse arrivano dal livello piú elevato. In altre parole, lo « spopolamento », o diseccitazione, delle particelle verso il livello laser inferiore deve essere della stessa entità dell'eccitazione di altre particelle dallo stato fondamentale verso il livello laser superiore, poiché una particella che ha già contribuito all'emissione laser deve ritornare allo stato fondamentale prima di essere nuovamente eccitata a un livello superiore e generare un incremento di potenza laser.

L'energia spesa dalle particelle nel ritorno verso lo stato fondamentale dai livelli laser inferiori non contribuisce all'emissione di potenza. Di conseguenza, per ogni particella che compie una transizione laser, viene persa una certa quantità di energia. Questo fatto suggerisce un sistema molto semplice per gudicare l'efficienza di un particolare tipo di laser. La quantità di energia persa da una particella per ritornare dai livelli laser inferiori verso lo stato fondamentale è uguale alla differenza fra l'energia necessaria per eccitare la particella allo stato laser superiore e la energia del fotone di luce che viene emesso quando la particella compie la transizione dai livelli laser superiori a quelli inferiori. Ne consegue che il rapporto di queste due quantità, energia emessa divisa per energia di eccitazione, costituisce una misura dell'efficienza operativa di un dato laser. La situazione nella quale a ogni particella eccitata a un livello laser superiore corrisponde l'emissione di un fotone di radiazione laser è naturalmente ideale; cosí facendo si considerano trascurabili tutte le altre transizioni che avvengono tra il livello laser superiore e gli altri livelli energetici inferiori. Di conseguenza il rapporto fra l'energia del fotone emesso e l'energia di eccitazione costituisce l'efficienza massima assoluta (o, come si è soliti dire, l'efficienza quantistica) del sistema laser.

In pratica, l'efficienza di un laser a gas è notevolmente piú bassa della sua efficienza quantistica, dato che non esiste un sistema perfetto per eccitare selettivamente le particelle del gas dallo stato fondamentale al livello laser superiore. Consideriamo il caso dell'eccitazione provocata dalle collisioni fra un atomo e un elettrone energetico durante la scarica in un gas. L'elettrone deve possedere una certa energia per eccitare l'atomo al livello laser superiore. In una scarica, gli elettroni non hanno però tutti la stessa energia cinetica, e anzi le loro energie sono distribuite con continuità su un grande intervallo di energie cinetiche. Di conseguenza un elettrone può eccitare gli atomi non solamente al livello laser superiore, ma anche ad altri livelli (piú alti o piú bassi del livello laser superiore), dai quali questi atomi non possono contribuire all'emissione laser. Il risultato è che solo una frazione della potenza elettrica fornita per produrre la scarica viene utilizzata per l'eccitazione degli atomi al livello laser superiore. Se definiamo l'efficienza operativa di un laser come il rapporto fra la potenza del fascio laser e la potenza fornita per la scarica elettrica, allora l'efficienza operativa sarà sempre piú bassa dell'efficienza quantistica. Piú un laser si avvicina a un sistema ideale, dal punto di vista della selettività del meccanismo di eccitazione, piú l'efficienza operativa sarà pros-

La molecola di anidride carbonica (*a*) ha una configurazione lineare e simmetrica, e possiede tre gradi di libertà vibrazionali. Nell'allungamento simmetrico lineare (*b*) gli atomi della molecola vibrano simmetricamente lungo l'asse internucleare. Nella vibrazione oscillatoria (*c*) il movimento degli atomi avviene perpendicolarmente all'asse internucleare. Nell'allungamento asimmetrico lineare (*d*) gli atomi vibrano lungo l'asse internucleare in modo asimmetrico. Lo stato vibrazionale della molecola è quindi descritto da tre numeri quantici ν_1, ν_2 e ν_3, e si rappresenta normalmente nella forma (v_1, v_2, v_3), essendo v_1 il numero di quanti vibrazionali simmetrico-lineari, v_2 il numero di quanti oscillatori e v_3, infine, il numero di quanti vibrazionali asimmetrico-lineari.

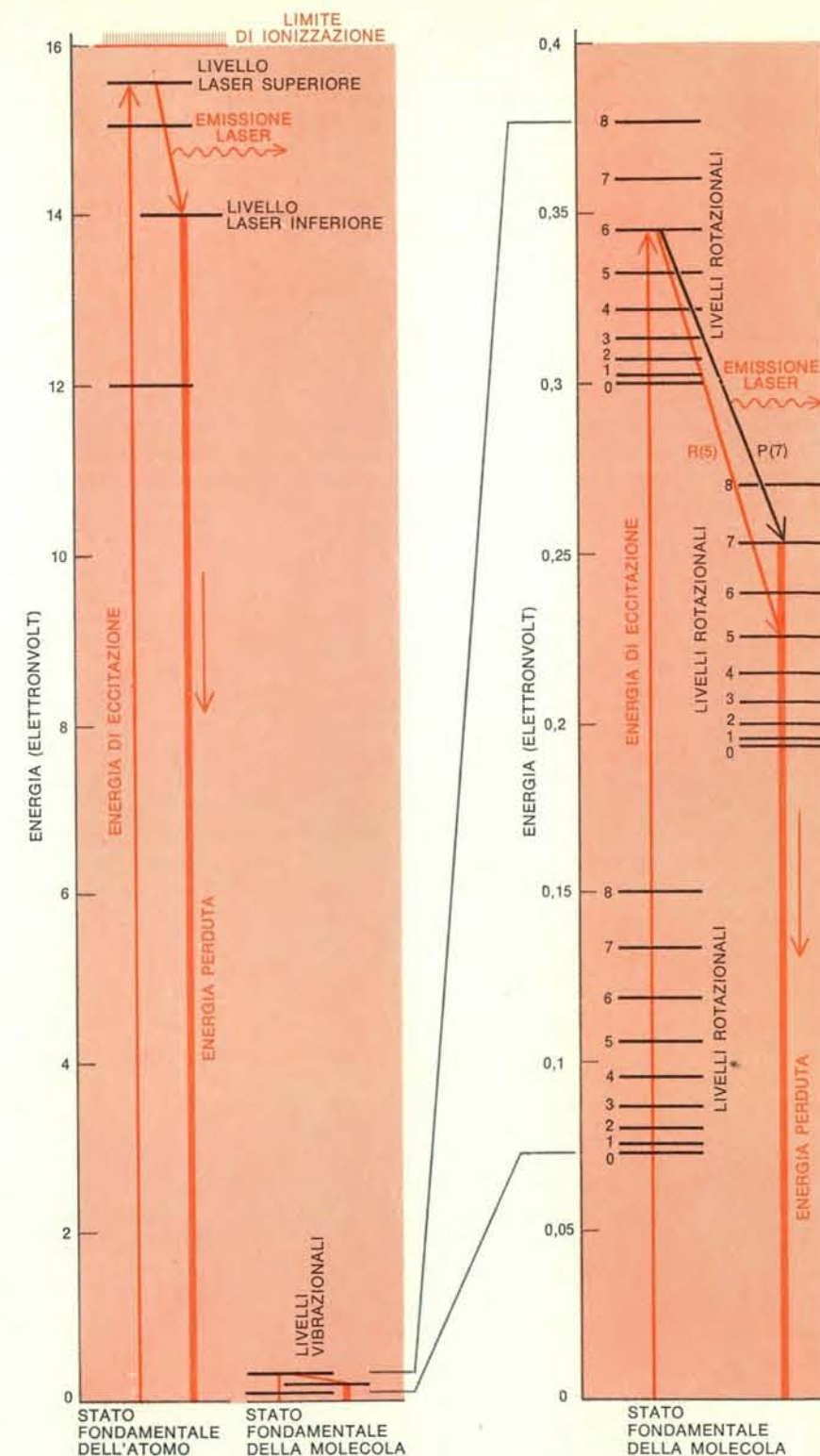
Il laser ad anidride carbonica di 54 m del Redstone Arsenal a Huntsville è in grado di emettere una radiazione coerente continua avente una potenza superiore a 2,5 kW.

sima all'efficienza quantistica. Oppure, da un altro punto di vista, un'elevata efficienza quantistica, unita a un meccanismo di eccitazione selettivo, costituisce la condizione per ottenere una elevata efficienza operativa in un sistema laser funzionale.

Il primo laser a gas fu costruito nel 1961 presso i Bell Telephone Laboratories. Esso sfruttava la transizione fra due stati eccitati del neo atomico, dando origine a una forte oscillazione laser con una lunghezza d'onda di 1,15 micron. Da allora l'effetto laser è stato ottenuto impiegando pressoché tutti gli elementi, coprendo un intervallo di lunghezze d'onda compreso fra 0,2 micron (2000 angstrom) nella regione dell'ultravioletto e 133 micron nella regione dell'infrarosso.

Gli spettri dei livelli energetici dei gas molecolari sono di gran lunga piú complicati di quelli dei gas atomici. Oltre ai comuni livelli energetici elettronici, una molecola può anche possedere livelli energetici dovuti ai suoi moti vibrazionali e rotazionali (*si veda la figura nella pagina a fronte*). Pertanto per una data configurazione elettronica di una molecola biatomica (formata cioè da 2 atomi), esistono diversi livelli energetici vibrazionali ugualmente spaziati, e per ognuno di tali livelli vibrazionali esiste un certo numero di livelli rotazionali. La spaziatura dei livelli energetici elettronici delle molecole è paragonabile a quella degli atomi, mentre la spaziatura dei livelli vibrazionali e rotazionali è notevolmente minore, precisamente di un fattore variabile fra 20 e 500. L'andamento dei livelli energetici di un gas molecolare è quindi estremamente complicato.

La prima oscillazione laser con gas molecolari venne ottenuta sfruttando le transizioni elettroniche di un certo numero di gas biatomici. Naturalmente è possibile avere anche transizioni fra due livelli vibrazionali diversi, appartenenti però allo stesso livello elettronico della molecola. Queste transizioni avvengono fra due livelli rotazionali appartenenti a due diversi livelli vibrazionali. Inoltre, a causa della crescente spaziatura dei livelli rotazionali di un livello vibrazionale, e a causa di una regola di selezione quantomeccanica, che permette nel caso piú semplice solo quelle transizioni che avvengono con una variazione di momento angolare rotazionale pari a $\pm h/2\pi$ (h è la costante di Planck), queste transizioni fra due livelli vibrazionali si manifestano come una banda di frequenze rotovibrazionali (*si veda la figura a pagina 84*). Il centro di questa banda corrisponde alla separazione esistente fra i livelli vibrazionali in assenza di energia rotazionale; le transizioni relative alle lunghezze d'onda maggiori corrispondono a variazioni del momento angolare di $+h/2\pi$ e sono chiamate transizioni del ramo P, mentre quelle relative alle lunghezze d'onda minori, dovute a una variazione del momento angolare di $-h/2\pi$, sono chiamate transizioni del ramo R. Come mostra la figura, il ramo P e il ramo R hanno la stessa larghezza. Queste transizioni rotovibrazionali, che si manifestano nell'emissione infrarossa, costituiscono il principio su cui si basa il funzionamento di tutti i laser a gas molecolari ad alta potenza oggi esistenti.

Vorrei spiegare ora come giungemmo, alcuni anni fa presso i Bell Telephone Laboratories, alla costruzione del primo laser a gas molecolare, rotovibrazionale, a emissione continua. Durante i nostri esperimenti riguardanti l'effetto laser nei gas atomici, capimmo che, se il nostro scopo era quello di ottenere una emissione ad alta potenza nella regione dell'infrarosso, cioè con lunghezze d'onda maggiori di alcuni micron, i gas atomici si discostavano molto dal sistema ideale. Ciò era principalmente dovuto al fatto che la maggior parte dei gas atomici presentano livelli elettronici (fra i quali si manifestano le transizioni infrarosse) molto prossimi ai limiti di ionizzazione atomica, cioè molto distanti dallo stato fondamentale dell'atomo. Di conseguenza, per ottenere l'effetto laser, l'atomo doveva essere eccitato a livelli energetici molto alti, e questo portava come conseguenza l'emissione di un fotone con energia relativamente bassa. L'impiego di gas atomici comportava quindi una bassa efficienza quantistica e un'ancor piú bassa efficienza operativa.

Questo sistema ha un ulteriore aspetto negativo. In prossimità del limite di ionizzazione di un elemento, i livelli corrispondenti a diverse configurazioni elettroniche sono raggruppati in un intervallo di energia molto stretto; di conseguenza, l'eccitazione prodotta dagli urti con gli elettroni risulta altamente non selettiva e la densità di popolazione degli atomi nel livello laser superiore risulta molto bassa. Questo fatto limita notevolmente l'emissione di potenza e origina un ulteriore abbassamento della efficienza operativa, dato che una notevole frazione degli elettroni in grado di eccitare gli atomi al livello laser su-

Confronto fra i livelli energetici di un atomo e quelli di una molecola. In un atomo i livelli energetici elettronici fra i quali possono avvenire transizioni che portano a emissione di radiazione infrarossa sono quelli situati in prossimità del limite di ionizzazione atomica, quindi molto lontani dallo stato fondamentale dell'atomo. In conseguenza a ciò l'atomo deve essere eccitato ad altissime energie perché si possa avere l'effetto laser, che si manifesta con l'emissione di un fotone di energia relativamente bassa. L'impiego di gas atomici comporta quindi un grande dispendio di energia e è associato a una bassa efficienza quantistica. In una molecola, al contrario, i livelli vibrazionali dello stato elettronico fondamentale sono molto prossimi al livello fondamentale della molecola; l'energia del fotone costituisce quindi una notevole frazione dell'energia totale richiesta per eccitare la molecola dallo stato fondamentale al livello laser superiore. Si ottiene cosí un'efficienza quantistica piú elevata. Nella figura a destra si può notare come i livelli vibrazionali dello stato fondamentale della molecola siano a loro volta costituiti da livelli energetici rotazionali dovuti alla rotazione molecolare. Il numero a fianco di ogni livello indica il valore del corrispondente momento angolare rotazionale in unità $h/2\pi$. Nel diagramma sono indicate due fra le transizioni infrarosse permesse, che interessano livelli rotazionali appartenuti a due diversi livelli vibrazionali.

periore vengono persi, poiché eccitano gli atomi anche ad altri livelli prossimi al livello laser. Un laser a gas atomico, operante a una lunghezza d'onda di circa 10 micron, ha solamente una potenza di alcuni milliwatt, e un'efficienza operativa dello 0,001 per cento circa.

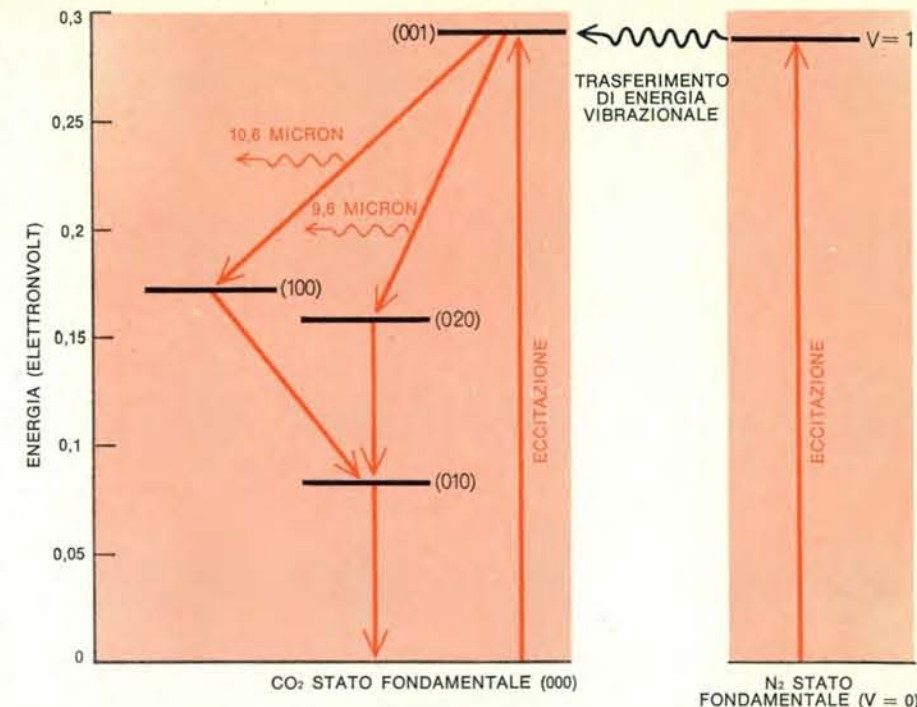
La situazione cambia completamente quando si opera con le molecole; i livelli rotovibrazionali appartenenti allo stato elettronico fondamentale della molecola sono l'ideale per un sistema laser efficiente e ad alta potenza, operante nella regione dell'infrarosso. I livelli vibrazionali dello stato elettronico fondamentale sono molto prossimi al livello fondamentale della molecola, e di conseguenza l'energia del fotone laser rappresenta una notevole frazione dell'energia totale necessaria per eccitare la molecola dallo stato fondamentale al livello laser superiore. Si ottiene cosí un'efficienza quantistica molto piú elevata di quella ottenibile con i laser a gas atomico operanti nell'infrarosso. Inoltre, poiché i livelli vibrazionali sono assai prossimi allo stato fondamentale della molecola, quasi tutti gli elettroni presenti nel processo di scarica partecipano all'eccitazione. Questo fatto assicura un'elevata efficienza operativa e una notevole emissione di potenza, dato che è possibile ottenere una alta densità di popolazione delle molecole nel livello laser superiore.

Prendendo lo spunto proprio da queste considerazioni decidemmo di vedere se era possibile impiegare per l'effetto laser le transizioni rotovibrazionali dello stato elettronico fondamentale dell'anidride carbonica. Sembrava che le molecole biatomiche fossero meno adatte per l'oscillazione laser a onda continua, dato che queste molecole biatomiche, eccitate a diversi livelli vibrazionali dello stato elettronico fondamentale, possiedono vite medie poco favorevoli. La nostra scelta si orientò sull'anidride carbonica per due motivi principali: primo, perché è una delle piú semplici molecole triatomiche; secondo, perché eravamo in possesso di un grande numero di informazioni spettroscopiche relative alle sue transizioni rotovibrazionali. La molecola dell'anidride carbonica è una molecola lineare a configurazione simmetrica e possiede tre gradi di libertà vibrazionali (*si veda la figura a pagina 81*). Il primo grado di libertà consente agli atomi che costituiscono la molecola di vibrare simmetricamente lungo l'asse internucleare. Questo tipo di vibrazione è detto « allungamento simmetrico lineare » e si denota con ν_1. Il secondo tipo di vibrazione simmetrica permette l'oscillazione degli atomi in direzione perpendicolare all'asse internucleare. Questa vibrazione, detta « oscillatoria », è indicata con ν_2. L'ultima possibilità, chiamata « allungamento asimmetrico lineare » e contrassegnata con il simbolo ν_3, permette la vibrazione asimmetrica lungo l'asse internucleare. Secondo le regole della meccanica quantistica le energie di queste vibrazioni sono quantizzate e sono l'una diversa dall'altra.

In prima approssimazione si può considerare che questi tre modi di vibrazione siano tra loro indipendenti. Di conseguenza, la molecola di anidride carbonica può essere eccitata in modo che la sua vibrazione effettiva sia una qualsiasi combinazione lineare di questi tre tipi fondamentali di vibrazione. Per questo lo stato vibrazionale della molecola deve essere descritto da tre numeri quantici, v_1, v_2 e v_3, che rappresentano i quanti dei diversi modi di vibrazione ν_1, ν_2 e ν_3 secondo i quali la molecola è eccitata. La descrizione di un determinato livello vibrazionale sarà di conseguenza espressa da un'espressione del tipo ($v_1 v_2 v_3$) dove v_1 è il numero di quanti vibrazionali simmetrico-lineari, v_2 è il numero di quanti vibrazionali simmetrico-oscillatori e v_3 il numero di quanti vibrazionali asimmetrico-lineari. Nel diagramma dei livelli energetici di alcuni stati vibrazionali dell'anidride carbonica (*si veda la figura a destra*) la struttura rotazionale di ciascuno dei livelli vibrazionali è stata esclusa per non complicare troppo il diagramma stesso. I livelli rotazionali sono molto meno spaziati di quanto non lo siano i livelli vibrazionali. I diversi livelli vibrazionali, caratterizzati da combinazioni diverse dei quanti dei modi ν_1, ν_2, ν_3, sono spaziati fra loro in maniera quasi uguale, ma nel diagramma sono indicati solamente quelli piú bassi (cioè quelli caratterizzati soltanto da uno o due quanti di energia vibrazionale). Per vari motivi, fra cui la vita media delle molecole dell'anidride carbonica nei diversi stati e la probabilità che l'impatto con un elettrone porti la molecola dallo stato fondamentale a uno stato eccitato, il livello contrassegnato con 001 è utilizzabile come livello laser superiore e i livelli 100 e 020 come livelli laser inferiori. Le molecole che giungono a questi livelli decadono allo stato fondamentale, attraverso lo stato 010, con transizioni a carattere radiativo e con transizioni indotte per collisione. Le transizioni rotovibrazionali 001 → 100 generano una radiazione infrarossa di lunghezza d'onda prossima a 10,6 micron, le transizioni 001 → 020 una radiazione infrarossa prossima a 9,6 micron. L'efficienza quantistica per la transizione laser 001 → 100 è prossima al 40 per cento, la transizione 001 → 020 ha un'efficienza prossima al 45 per cento. Queste elevate efficienze, e la possibilità di produrre eccitazioni selettive su livelli vicini al livello fondamentale, hanno invogliato a studiare questo sistema e ci hanno permesso in seguito di ottenere efficienze pratiche dell'ordine del 20-30 per cento.

L'oscillazione laser dovuta alle transizioni fra due livelli di energia rotazionale appartenenti a due diversi livelli vibrazionali dello stesso livello elettronico dell'anidride carbonica comporta l'emissione di radiazione infrarossa di un certo numero di lunghezze d'onda. Questo insieme di frequenze dà origine a quella che si chiama banda rotovibrazionale. La curva schematizza la posizione delle transizioni osservabili tramite l'assorbimento spettroscopico attraverso anidride carbonica gassosa non eccitata. Il centro di questa banda corrisponde alla separazione esistente fra i livelli vibrazionali in assenza di energia rotazionale; le transizioni relative alle lunghezze d'onda maggiori corrispondono a variazioni del momento angolare di $+h/2\pi$ e sono chiamate transizioni del ramo *P*, mentre quelle relative alle lunghezze d'onda minori, dovute a una variazione del momento angolare di $-h/2\pi$, sono chiamate transizioni del ramo *R* (*h* è la costante di Planck). La radiazione prodotta è di 10,6 micron.

Nei nostri primi esperimenti, il tubo laser era riempito con anidride carbonica pura, a una pressione di circa un millimetro di mercurio. La scarica elettrica era prodotta da una corrente continua ad alta tensione applicata attraverso una sezione del tubo. Durante tale scarica si producono un grande numero di collisioni fra gli elettroni accelerati e le molecole di anidride carbonica. Alcuni di questi elettroni producono la dissociazione molecolare della anidride carbonica, cioè suddividono la molecola in atomi di carbonio e di ossigeno. La soglia energetica per questo fenomeno è tuttavia piuttosto elevata e il numero di elettroni che posseggono una cosí grande energia cinetica è molto basso. Gli elettroni che posseggono il minimo di energia cinetica, che sono di gran lunga superiori numericamente agli elettroni di alta energia, producono l'eccitazione ai vari livelli vibrazionali delle molecole di anidride carbonica. Normalmente avviene che gli elettroni eccitino la molecola di anidride carbonica in preferenza al livello $00v_3$, cioè ai livelli ugualmente spaziati dello stato ν_3.

È necessario ricordare che per il livello superiore dell'oscillazione laser, corrispondente a 10,6 micron di lunghezza d'onda, è $v_3 = 1$. Ci si potrebbe quindi chiedere se le molecole di anidride carbonica che vengono eccitate agli stati superiori a $00v_3$ (con v_3 maggiore di 1) non contribuiscono all'effetto laser, e di conseguenza non riducono l'efficienza e la potenza di emissione di questo sistema. Questo d'altra parte non dovrebbe essere vero dato che i livelli $00v_3$ sono ugualmente spaziati, e perciò una collisione fra una molecola $CO_2(00v_3)$ con una molecola $CO_2(000)$ dovrebbe avere come unica conseguenza un trasferimento di energia vibrazionale da una molecola eccitata a una molecola non eccitata. La molecola $CO_2(00v_3)$ perde un quanto di energia vibrazionale ν_3 e diviene una molecola $CO_2(00v_3\text{-}1)$, mentre la $CO_2(000)$ guadagna un quanto di energia diventando una molecola $CO_2(001)$, cioè, in altre parole, la molecola si porta al livello superiore dello stato laser (*si veda la figura in alto alle pagine 86 e 87*).

L'immissione di azoto in un laser ad anidride carbonica consente di eccitare selettivamente le molecole di anidride carbonica al livello laser superiore. La molecola di azoto, essendo biatomica, possiede due soli gradi di libertà vibrazionali; un solo numero quantico vibrazionale (v) è in grado di descrivere completamente i suoi livelli energetici vibrazionali. Le molecole di azoto possono essere efficientemente eccitate, per collisione elettronica nella scarica prodotta in azoto a bassa pressione, dal livello $v = 0$ al livello $v = 1$. Poiché l'energia di eccitazione della molecola $N_2(v = 1)$ è all'incirca uguale all'energia di eccitazione della molecola $CO_2(001)$, è possibile il trasferimento dell'energia vibrazionale dalla molecola di azoto a quella di anidride carbonica, in seguito a collisioni fra molecole $N_2(v = 1)$ e molecole $CO_2(000)$. A causa di tali collisioni la molecola di azoto ritorna dal livello $v = 1$ al livello fondamentale perdendo un quanto della sua energia vibrazionale, mentre una molecola di anidride carbonica si eccita dallo stato fondamentale allo stato 001. Essa, in seguito, può decadere radiativamente sia al livello 100 sia al livello 020, emettendo radiazioni infrarosse di 10,6 e 9,6 micron.

Questo processo è risonante, nel senso che esiste una ridistribuzione della energia delle molecole eccitate senza alcuna perdita dell'energia totale interna sotto forma di energia cinetica o termica. Questo significa che è molto alta l'efficienza di conversione delle molecole $CO_2(00v_3)$ in molecole $CO_2(001)$ senza perdite di energia. In pratica perciò si potrebbero eccitare le molecole di anidride carbonica al livello laser superiore con alta efficienza, servendosi delle collisioni con gli elettroni in una scarica nel gas.

Le molecole $CO_2(001)$ potrebbero ora, per esempio, emettere un fotone laser con lunghezza d'onda di 10,6 micron, e cadere nello stato 100, dal quale potrebbero poi ritornare allo stato fondamentale prima che la molecola possa essere riutilizzata per la produzione di un nuovo fotone laser. Le molecole che si trovano al livello laser inferiore sono normalmente diseccitate in seguito a collisioni con altre molecole. Ancora una volta i trasferimenti risonanti di energia vibrazionale assumono un ruolo molto importante. Il livello laser inferiore possiede un'energia che è circa due volte l'energia necessaria per eccitare la molecola di anidride carbonica allo stato vibrazionale 010. Perciò in una collisione che interessa una molecola $CO_2(100)$ o una molecola $CO_2(020)$ con una molecola $CO_2(000)$, l'energia vibrazionale si distribuisce con alta efficienza fra le due molecole eccitandole entrambe al livello 010 (*si veda la figura in basso alle pagine 86 e 87*).

A causa della natura risonante di queste collisioni la diseccitazione vibrazionale del livello laser inferiore è senza dubbio efficiente. Il processo di diseccitazione non è comunque del tutto completo. Le molecole $CO_2(010)$ devono ancora essere diseccitate allo stato fondamentale prima che possano riprendere parte al processo di emissione laser. Anche la diseccitazione delle molecole $CO_2(010)$ è governata da un processo di urto, però, in questo caso, il processo non è risonante e l'energia di queste molecole deve essere trasformata in energia cinetica. Queste collisioni

possono aver luogo con altre molecole di CO_2, con altre particelle gassose presenti o anche con le pareti del tubo.

A causa della natura non risonante del processo di conversione della energia vibrazionale in energia cinetica, la diseccitazione delle molecole $CO_2(010)$ può essere lenta e costituire una « strozzatura » per i successivi cicli di eccitazione e diseccitazione, riducendo di conseguenza l'efficienza e l'emissione di potenza. Ho avuto occasione di verificare che anche usando anidride carbonica pura il meccanismo di diseccitazione è sufficientemente rapido da permettere forti oscillazioni laser relative a transizioni rotovibrazionali fra i livelli 001 → 100 e 001 → 020 che corrispondono a emissioni aventi rispettivamente lunghezze d'onda di 10,6 e 9,6 micron. Si è trovato che a causa della maggiore probabilità di emissione le transizioni relative a 10,6 micron risultano piú intense di circa un fattore 10 di quelle relative a 9,6 micron. Nelle successive considerazioni che sono presentate in questo articolo, ci riferiremo solamente alla transizione che porta all'emissione di radiazioni di lunghezze di onda di 10,6 micron.

È ben chiaro che l'eccitazione per urto con un elettrone, che avviene in conseguenza di una scarica in anidride carbonica pura, non è in grado di produrre quell'eccitazione ad alta selettività delle molecole al livello laser superiore che è necessaria per ottenere un'efficienza di lavoro prossima alla efficienza quantistica del sistema. Il motivo è che l'elettrone può anche eccitare le molecole di anidride carbonica ad altri livelli, diversi dal livello $00v_3$, e questo fatto provoca una riduzione sia dell'efficienza sia della potenza di emissione. Per ottenere un'elevata efficienza è necessario che le molecole di anidride carbonica siano eccitate selettivamente al livello laser superiore. Questa eccitazione selettiva si manifesta quando all'anidride carbonica, usata per il funzionamento del laser, si aggiunge azoto.

L'utilità dell'azoto può essere spiegata riferendosi al diagramma dei livelli energetici, in particolar modo ai livelli vibrazionali piú bassi dello stato fondamentale dell'azoto molecolare (*si veda la figura a pagina 85*). Poiché la molecola di azoto è biatomica, essa possiede un solo grado di libertà vibrazionale; i suoi livelli energetici vibrazionali sono espressi da quanti energetici derivanti da vibrazioni che si manifestano solamente lungo l'asse internucleare. Di conseguenza un solo numero quantico descrive completamente i livelli vibrazionali della molecola di azoto. Poiché l'azoto è costituito da molecole biatomiche e omonucleari, l'azoto molecolare eccitato a diversi livelli vibrazionali dello stato elettronico fondamentale non può decadere irradiando o per collisioni, per cui questi stati hanno vita estremamente lunga.

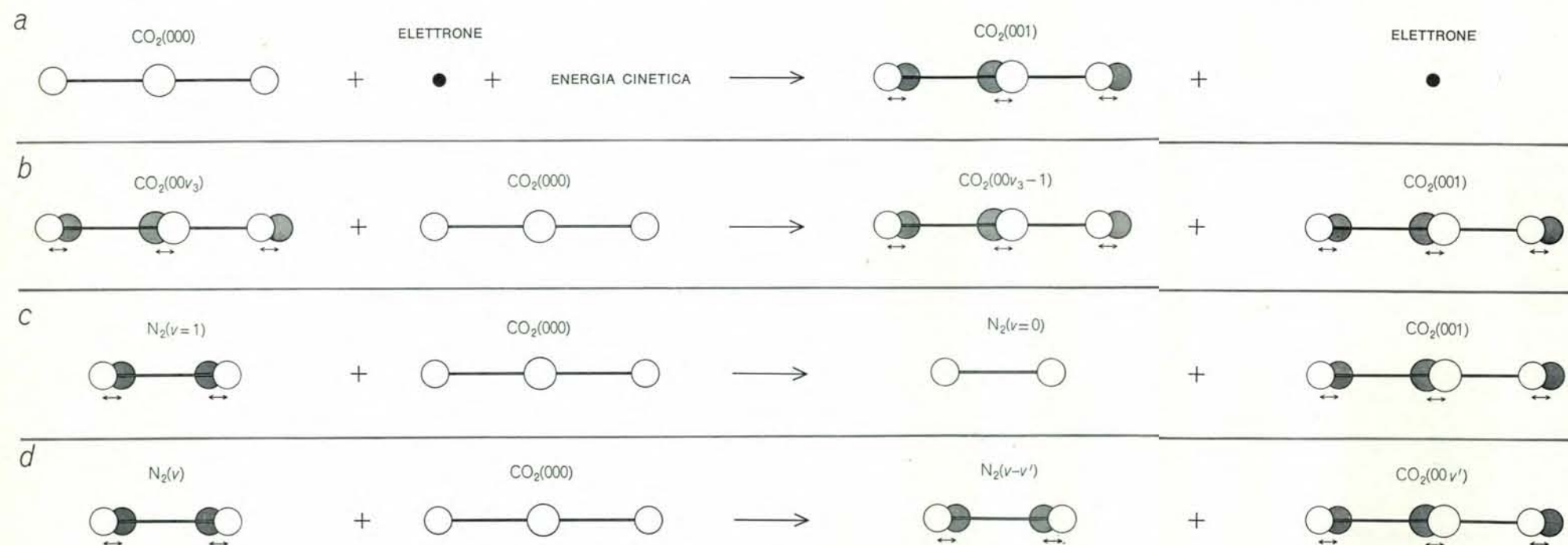
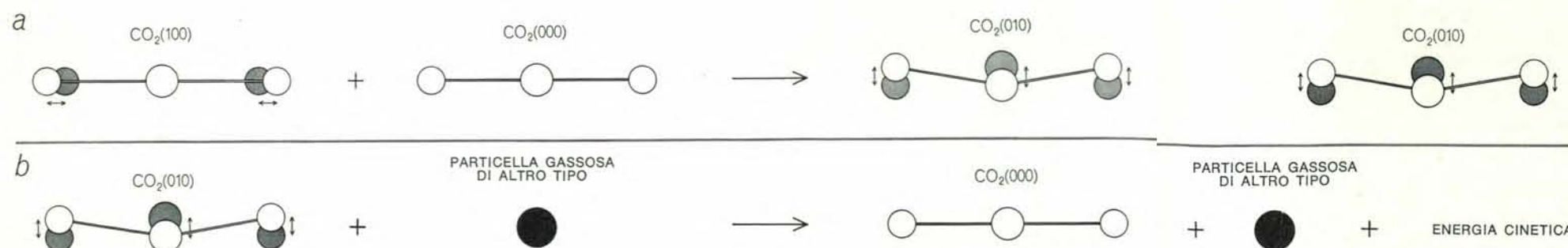
Le molecole di azoto sono eccitate dal livello $v = 0$ a vari altri livelli vibrazionali piú elevati, soprattutto per urto con elettroni; possono inoltre venire eccitate in seguito a diseccitazione da livelli elettronici piú elevati o alla ricombinazione di atomi di azoto dissociati. In una scarica in azoto a bassa pressione è possibile eccitare circa il 30 per cento delle molecole al livello $v = 1$. Poiché l'energia di eccitazione della molecola $N_2(v = 1)$ è circa uguale all'energia di eccitazione della molecola $CO_2(001)$, ci si potrebbe aspettare un trasferimento di energia vibrazionale dall'azoto all'anidride carbonica, per collisione fra la molecola $N_2(v = 1)$ e la molecola $CO_2(000)$. In seguito a questa collisione, la molecola di azoto ritorna dal livello $v = 1$ al suo stato fondamentale, liberando un quanto della sua energia vibrazionale, mentre la molecola di anidride carbonica viene eccitata dal suo stato fondamentale al livello 001. A causa della natura risonante di questo processo di collisione, l'eccitazione selettiva delle molecole di anidride carbonica al livello laser superiore dovrebbe essere molto efficiente.

Inoltre, i livelli vibrazionali piú elevati della molecola di azoto hanno una spaziatura quasi uguale a quella dei livelli di $CO(00v_3)$. Pertanto, nelle collisioni che interessano le molecole $N_2(v)$ e $CO_2(000)$ si manifestano dei trasferimenti efficienti di energia vibrazionale nei quali le molecole $N_2(v)$ perdono v' quanti di energia vibrazionale, diseccitandosi al livello $N_2(v-v')$ a favore delle molecole $CO_2(000)$ che acquistando v' quanti di energia vibrazionale si eccitano selettivamente al livello CO_2 $(00v_3 = v')$. Poiché la spaziatura dei livelli energetici degli stati di $N_2(v)$ è quasi uguale a quella degli stati di $CO_2(00v_3)$, queste collisioni coinvolgono trasferimenti risonanti di energia vibrazionale e il processo ha un'elevata efficienza. Le molecole $CO_2(00v_3 = = v')$ si trasformano in $CO_2(001)$ (cioè in molecole nel livello laser superiore) in seguito alle collisioni risonanti precedentemente trattate. In definitiva si ottiene un'eccitazione selettiva efficiente di molecole di anidride carbonica al livello laser superiore, e ci si può quindi aspettare che un laser ad anidride carbonica e azoto sia piú efficiente e abbia un'emissione di potenza superiore a quella di un laser riempito solamente con anidride carbonica pura.

I primi esperimenti fatti per verificare questa ipotesi vennero eseguiti con l'apparecchiatura schematizzata a pagina 90. In questa apparecchiatura il gas fluiva con continuità. Nella regione di interazione, dove deve avvenire l'azione laser, non veniva prodotta la scarica elettrica. L'azoto veniva immesso e fatto passare nella regione di eccitazione, nella quale veniva prodotta la scarica per mezzo di un campo elettrico oscillante o applicando un'alta tensione continua. Durante il passaggio nella regione di scarica, le molecole di azoto venivano eccitate ai diversi livelli vibrazionali dello stato elettronico fondamentale.

Poiché il sistema funzionava a flusso continuo, le molecole di azoto che erano passate nella regione di scarica venivano aspirate nella regione d'interazione, in un intervallo di tempo breve rispetto alla vita media dello stato vibrazionale delle molecole stesse. Di conseguenza l'azoto che entrava nella regione d'interazione possedeva un notevole numero di molecole che si trovavano ancora, e che restavano, nello stato vibrazionale eccitato dello stato elettronico fondamentale. L'anidride carbonica, introdotta attraverso un'altra entrata, si mescolava con l'azoto. Come è stato precedentemente detto, la energia vibrazionale si trasferiva dallo azoto all'anidride carbonica, dato che le collisioni che si verificavano fra le molecole interessavano molecole di azoto in uno stato vibrazionale e molecole di anidride carbonica nello stato fondamentale. Le molecole di anidride carbonica venivano in questo modo eccitate selettivamente allo stato laser superiore. Si fa notare che non vi era alcuna altra forma di eccitazione per l'anidride carbonica nella regione di interazione.

Con questo sistema si possono pertanto ottenere forti oscillazioni laser sulle transizioni rotovibrazionali della anidride carbonica, anche senza produrre la scarica nella regione di interazione. Dopo che le molecole di anidride carbonica avevano portato il loro contributo all'oscillazione laser, esse venivano aspirate dal sistema a flusso continuo, in modo da permettere l'ingresso di nuove molecole di azoto appena eccitate e di nuove molecole di anidride carbonica, che davano cosí il loro contributo alla continuità dell'oscillazione laser. La potenza dell'oscillazione laser costituisce una prova dell'efficienza del sistema che impiega le molecole di azoto eccitate a un livello vibrazionale per l'eccitazione delle molecole di anidride carbonica al livello laser superiore. Mescolando insieme azoto e anidride carbonica in un tubo laser e facendo avvenire la scarica nella regione laser stessa, vi sono ottenute efficienze di conversione superiori al 5 %.

Con un laser ad anidride carbonica

In questa figura sono schematizzati i meccanismi di eccitazione che portano le molecole di anidride carbonica al livello laser superiore (in questo caso il livello 001). Se in una scarica elettrica un elettrone energetico colpisce la molecola di anidride carbonica non eccitata (cioè al livello 000) è possibile che questa si ecciti direttamente al livello 001 (*a*). In un altro caso questa collisione può produrre l'eccitazione dal livello 000 al livello $00v_3$, dove v_3 indica un'oscillazione asimmetrica lineare con un numero di quanti di energia vibrazionale superiore a uno; in questo caso successive collisioni con molecole non eccitate possono provocare trasferimenti di singoli quanti di energia vibrazionale verso molecole non eccitate, che vengono cosí a raggiungere il livello 001 (*b*). Nel laser ad anidride carbonica e azoto, l'urto fra molecole di azoto eccitate a un livello vibrazionale e molecole non eccitate di anidride carbonica può portare queste ultime al livello 001, in seguito al trasferimento di un quanto di energia vibrazionale dalla molecola di azoto, che può avere, prima che si verifichi la collisione, uno solo (*c*) o piú (*d*) quanti di energia vibrazionale.

I meccanismi di diseccitazione, producendo lo « spopolamento » dei livelli vibrazionali inferiori dell'anidride carbonica, possono aumentare la potenza del laser. Uno di questi meccanismi è la collisione di una molecola eccitata al livello 100 con una molecola 000 non eccitata, che lascia entrambe le molecole al livello 010 (*a*); un altro è la collisione di una molecola di anidride carbonica 010 con una particella di altro tipo (o con le pareti del tubo), che porta la molecola al livello fondamentale (*b*).

si potrebbe ottenere un aumento della potenza di emissione e dell'efficienza aumentando la probabilità di diseccitazione verso il livello laser inferiore, eliminando cioè la « strozzatura » del livello 010. All'inizio ho ricordato che la diseccitazione delle molecole CO_2(010) avviene in seguito alla conversione dell'energia da esse posseduta in energia cinetica in un processo di collisione con altre particelle. La velocità con la quale questo processo si manifesta dipende dalla natura delle particelle con le quali l'anidride carbonica collide. Per esempio, la stessa anidride carbonica, alla pressione di 1 millimetro di mercurio, presenta 100 diseccitazioni al secondo, mentre l'elio, alla stessa pressione, ne presenta 4000 (e le molecole d'acqua circa 100 000). Possiamo quindi disporre di un altro sistema per aumentare la potenza di uscita e l'efficienza di un laser ad azoto e anidride carbonica.

Si è scoperto che per ottenere una potenza di emissione continua e molto elevata, e un'alta efficienza, è necessario introdurre nel tubo di scarica ulteriori gas, come ossigeno, vapor acqueo, idrogeno e elio, i quali producono un aumento della potenza. L'aumento deve essere spiegato alla luce di due effetti: 1) aumento della diseccitazione del livello vibrazionale inferiore delle molecole di anidride carbonica e 2) aumento della quantità delle molecole di anidride carbonica eccitate al livello 001, sia in modo diretto, attraverso il processo d'urto con gli elettroni, sia in modo indiretto, in seguito all'aumento dei processi di eccitazione vibrazionale da parte delle molecole di azoto. Entrambi i processi che portano a un aumento dell'eccitazione delle molecole di anidride carbonica al livello laser superiore sono possibili se aumenta la densità degli elettroni nel processo di scarica e anche se varia la loro distribuzione energetica, in modo che risultino piú favorite sia l'eccitazione delle molecole di anidride carbonica direttamente al livello 001, sia una piú rapida produzione di molecole $N_2(v)$.

L'impiego dell'elio sembra essere piú adatto per entrambi gli scopi, e per questo motivo questo gas viene impiegato con maggiore frequenza come terzo gas di riempimento. Il vapor acqueo e l'idrogeno presentano un'utilità pratica solamente nei confronti del primo effetto. L'impiego dell'ossido di carbonio presenta una certa utilità sia dal punto di vista dell'eccitazione sia della diseccitazione. Impiegando una miscela costituita da anidride carbonica e azoto alla pressione di tre millimetri di mercurio, ed elio alla pressione di circa 20 millimetri di mercurio, si è ottenuta l'emissione di un'onda continua, con una lunghezza d'onda di 10,6 micron, della potenza di circa 80 watt per metro di lunghezza del tubo di scarica. L'efficienza operativa è risultata superiore al 20 %.

La distribuzione di Boltzmann della densità di affollamento dei livelli energetici rotazionali, relativa al livello vibrazionale 001 dell'anidride carbonica, è dovuta al fatto che una molecola, durante il periodo nel quale si trova in un determinato livello vibrazionale, è soggetta a un grande numero di collisioni di termalizzazione rotazionale; in un secondo la molecola compie circa 10 milioni di transizioni da un livello rotazionale a un altro. La densità di affollamento dei livelli rotazionali riportata sull'asse orizzontale si riferisce a una temperatura di circa 400 gradi Kelvin, ossia circa 123 gradi centigradi.

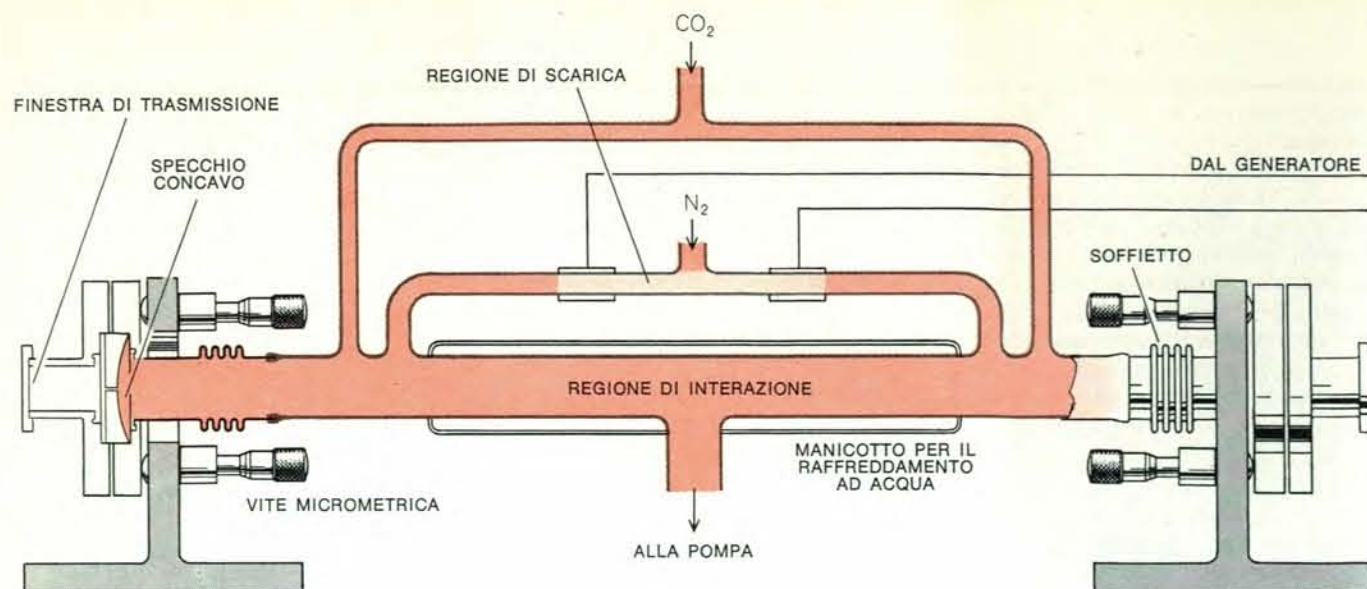
Attualmente, la maggior parte dei laser ad anidride carbonica ad alta potenza funzionano con gas che fluiscono lentamente nel tubo laser. Alcuni fra i nostri primi esperimenti eseguiti ai Bell Laboratories, ed esperimenti piú recenti compiuti al Philips Research Laboratory in Olanda, hanno mostrato che è possibile costruire laser ad anidride carbonica a ciclo chiuso, se si pone una grande attenzione nella preparazione dei contenitori e se viene usata l'esatta miscela di gas. Questi laser sono in grado di sviluppare una potenza analoga a quella sviluppata da un sistema a flusso gassoso, e hanno un'efficienza dello stesso ordine di grandezza.

I tipi piú comuni di laser ad anidride carbonica sono lunghi circa due metri e possono generare un'onda laser continua avente una potenza di circa 150 watt. Non esiste nessuna limitazione che impedisca la costruzione di un laser molto lungo, capace di generare un'emissione di potenza molto elevata, dato che la potenza emessa aumenta linearmente con la lunghezza del laser. In effetti, alcuni ricercatori della Raytheon Company hanno costruito un laser ad anidride carbonica lungo 180 metri, in grado di emettere una potenza superiore a 8,8 chilowatt. L'emissione di potenza del laser ha cosí raggiunto valori da fantascienza e gli effetti termici di queste emissioni sono senza dubbio terrificanti.

La disposizione sperimentale mostrata nella figura a pagina 90 è molto utile per studiare l'effetto laser nei gas molecolari instabili, cioè in quei gas che si dissociano facilmente sotto l'azione diretta della scarica e/o che richiedono un'eccitazione estremamente selettiva per un funzionamento laser continuo (per esempio i gas biatomici). L'utilità di questa apparecchiatura sta nel fatto che la scarica non viene prodotta nella regione laser e pertanto i soli livelli possibili per il gas attivo che deve essere eccitato sono quelli per i quali l'energia coincide con i livelli vibrazionali delle molecole di azoto eccitate. Si ottengono in questo modo oscillazioni laser continue, per mezzo delle transizioni rotovibrazionali dell'ossido di carbonio (CO) da cinque a sei micron, dell'ossido di azoto (N_2O) sulle transizioni 001 → 100 corrispondenti a 10,8 micron e del bisolfuro di carbonio (CS_2) sulle transizioni 021 → 120 corrispondenti a 11 micron.

Il laser a ossido di carbonio è particolarmente interessante da un punto di vista spettroscopico. Siamo ora in grado di osservare transizioni, nello stato elettronico fondamentale, fra livelli vibrazionali superiori a $v = 25$, transizioni che non sono mai state osservate in precedenza. L'ossido di azoto, che presenta notevoli somiglianze, nei suoi stati vibrazionali, con l'anidride carbonica, ha la capacità di produrre onde laser continue ad alta potenza, con un'efficienza paragonabile a quella di un sistema ad anidride carbonica: questo sistema non è stato tuttavia studiato come il precedente. Sembra, comunque, che sia possibile applicare in modo del tutto generale la tecnica dell'eccitazione selettiva per ottenere elevate emissioni di potenza e alte efficienze.

Fra le altre tecniche che permettono di ottenere eccitazioni vibrazionali delle molecole si possono includere le reazioni chimiche, il riscaldamento dei gas con fiamma o bruciatori e l'eccitazione ottica ottenuta per mezzo di radiazioni luminose provenienti da una scarica o dalla fiamma. Questi sistemi non sono ancora stati impiegati a fondo, tuttavia per essi esistono buone possibilità. Il fatto che l'eccitazione ottenuta mediante la scarica, nel laser ad anidride carbonica, sia in grado di fornire un'efficienza di conversione superiore al 20 per cento costituisce tuttavia un notevole ostacolo, dal punto di vista dell'applicazione pratica, per l'affermazione di nuove tecniche.

Gli effetti competitivi fra le possibili transizioni laser rotovibrazionali dell'anidride carbonica si manifestano con la predominanza di una ben determinata transizione del ramo *P*. Le curve rappresentano gli incrementi (o i decrementi) di alcune transizioni del ramo *P* (*curve colorate*) e del ramo *R* (*curve nere*) per una data banda vibrazionale. I valori numerici associati a ogni curva esprimono il rapporto fra la densità totale di affollamento nel livello vibrazionale laser superiore (001) e in quello inferiore (100).

Questo sistema a flusso continuo è stato impiegato per verificare l'ipotesi secondo cui un laser ad anidride carbonica e azoto è più efficiente di uno ad anidride carbonica pura. Con questo sistema sono state ottenute forti oscillazioni laser sulle transizioni rotovibrazionali dell'anidride carbonica senza applicare la scarica elettrica nella regione d'interazione, ma servendosi delle molecole di azoto, precedentemente eccitate a livelli vibrazionali, per eccitare selettivamente le molecole di anidride carbonica.

Finora ho descritto i meccanismi di eccitazione e diseccitazione che producono nei laser ad anidride carbonica un'emissione di potenza molto elevata; non ho tuttavia ancora detto niente a proposito dello spettro di emissione. Come è già stato detto, le transizioni fra due livelli vibrazionali si manifestano sotto forma di una banda che risulta costituita dai rami P e R, conseguenza dei livelli rotazionali poco spaziati appartenenti sia a livello vibrazionale laser superiore sia a quello inferiore. Ci si può a questo punto chiedere se questo fatto significa che l'emissione di potenza di un laser ad anidride carbonica avviene simultaneamente con un numero di frequenze corrispondenti alle transizioni dei rami R e P. Se l'emissione avvenisse con un numero discreto di frequenze, il fascio laser non sarebbe monocromatico nel vero senso della parola, e la sua utilità, in certe applicazioni, come per esempio nel campo delle comunicazioni, risulterebbe limitata. In realtà è possibile che l'emissione di potenza di un laser ad anidride carbonica avvenga su una sola transizione del ramo P, normalmente per $P(20)$, corrispondente a una lunghezza d'onda di 10,5915 micron, senza alcuna complicazione, a dispetto del fatto che la banda vibrazionale 001 → 100 contiene un numero discreto di transizioni possibili sui rami P e R. Tutto questo è molto chiaro se si considerano alcuni « effetti competitivi » molto fini che avvengono fra le transizioni dei rami P e R. Questo fenomeno aumenta in modo incredibile l'utilità di impiegare l'anidride carbonica nel laser.

È importante osservare che l'intervallo energetico esistente tra i diversi livelli vibrazionali è normalmente piú grande dell'energia cinetica delle molecole (che risulta dell'ordine di 0,025 elettronvolt a temperatura ambiente). La spaziatura energetica dei livelli rotazionali è, d'altro canto, inferiore a questa energia cinetica. Pertanto la densità di affollamento per un particolare livello rotazionale appartenente a un dato livello vibrazionale non è indipendente dalla densità degli altri livelli rotazionali, dato che ogni singola collisione dà origine a uno scambio di energia uguale alle spaziature dei livelli rotazionali. In conclusione la molecola può, molto frequentemente, saltare da un livello rotazionale a un altro. La frequenza di questi spostamenti (noti anche come termalizzazioni rotazionali) è, in un gas che si trovi alla pressione di funzionamento di un laser, superiore a 10 milioni di salti al secondo. Poiché la spaziatura dei livelli energetici vibrazionali è molto maggiore dell'energia cinetica delle molecole, l'effetto della termalizzazione vibrazionale è molto basso (circa 1000 spostamenti al secondo). La vita di un livello vibrazionale, includendo i fenomeni radiativi e di collisione, è di circa un millisecondo, mentre la vita media di termalizzazione rotazionale è considerevolmente piú corta, essendo di circa 10^{-7} secondi. Questo implica che, durante la sua vita in un livello vibrazionale, una molecola è soggetta a un elevatissimo numero di collisioni di termalizzazione rotazionale. Per questo motivo la probabilità di esistenza dei vari livelli rotazionali di una molecola, in un determinato livello vibrazionale, segue una distribuzione di Boltzmann (*si veda la figura a pagina 88*).

Alla luce di questa ulteriore condizione che interessa la densità di affollamento dei livelli rotazionali è possibile calcolare la maggiore (o minore) probabilità delle singole transizioni, relative ai rami R e P, per una data banda vibrazionale (*si veda la figura a pagina 89*). La serie di curve ottenute da tale calcolo porta alle seguenti utili conclusioni: 1) alcune transizioni del ramo P manifestano un incremento di probabilità anche quando la densità di affollamento vibrazionale totale relativa al livello laser inferiore supera quella del livello superiore. Questa situazione viene detta « inversione parziale », dato che le transizioni del ramo R non manifestano un simile incremento. 2) Quando la densità di affollamento vibrazionale totale del livello laser superiore supera quella del livello inferiore, sia le transizioni del ramo P che quelle del ramo R manifestano un incremento. Questa situazione viene detta « inversione completa ». 3) Anche nel caso dell'inversione completa, una transizione relativa al ramo R manifesta sempre un incremento minore di una transizione del ramo P proveniente dallo stesso livello rotazionale superiore.

Vediamo ora quali effetti pratici si possono osservare sulla base delle ultime conclusioni. Mentre è vero che si hanno simultaneamente incrementi per un grande numero di transizioni, l'esistenza di una distribuzione di Boltzmann comporta che la variazione della densità di affollamento di un livello rotazionale influenzi la densità di tutti i livelli rotazionali, in modo che sia mantenuta la distribuzione di Boltzmann. La transizione rotazionale con l'incremento piú elevato – in questo caso la transizione $P(22)$ – inizierà l'oscillazione per prima. Questa sarà la piú intensa transizione del ramo P, dato che le transizioni del ramo R hanno incrementi inferiori. Quando avviene un simile fenomeno aumenta la velocità alla quale le molecole vengono rimosse dal livello rotazionale $J = 21$, a causa dell'emissione stimolata sulla transizione $P(22)$. Inoltre, a causa della distribuzione di Boltzmann, si avrà un trasferimento di molecole dagli altri livelli rotazionali sul livello $J = 21$ e la densità di affollamento di tutti i livelli rotazionali diminuirà, anche se l'oscillazione laser su $P(22)$ porta via le molecole dal livello $J = 21$. Questo fatto costituisce un fortissimo effetto competitivo fra le possibili transizioni laser e per questo una sola transizione del ramo P risulta dominante.

A causa di questo effetto competitivo, che è il piú importante, l'emissione di potenza da parte di un laser ad anidride carbonica può avvenire su una singola transizione rotazionale della banda 001 → 100, e questo ci assicura che l'alta potenza emessa sia allo stesso tempo estremamente coerente ed estremamente monocromatica.

È possibile ottenere l'oscillazione anche sulle transizioni piú deboli dei rami P e R, purché esista un incremento sufficiente di probabilità e purché si introduca nella cavità laser un elemento selettore di lunghezze d'onda, come per esempio un reticolo o un prisma che interdica l'oscillazione delle transizioni piú forti. Impiegando un sistema selettore di lunghezze d'onda, a causa del forte effetto competitivo, qualunque sia la transizione scelta per l'oscillazione si ottiene sempre un'emissione di potenza quasi costante.

Inoltre, a causa della lunga vita media dei livelli vibrazionali, che nel laser ad anidride carbonica sono i responsabili del processo oscillatorio, è possibile, intercettando il cammino del raggio laser nel risuonatore e bloccando cosí l'oscillazione laser, immagazzinare energia nel mezzo nel quale avviene la scarica per circa un millisecondo. Se si toglie rapidamente il blocco, l'emissione laser può allora avvenire in forma impulsiva e l'impulso di potenza risulta circa 1000 volte superiore alla potenza media dell'onda continua emessa dal laser finora descritto. Questo tipo di funzionamento è chiamato a interruzione Q. Questa interruzione Q si ottiene molto piú facilmente sostituendo uno spec-

L'interruzione Q è un accorgimento tecnico impiegato per ottenere da un laser funzionante con continuità un'emissione impulsata. Questo accorgimento è ottenuto sostituendo uno degli specchi della cavità laser con uno specchio rotante. Il laser funziona ogni volta che lo specchio rotante (*a sinistra*) risulta parallelo allo specchio fisso che si trova sull'altra estremità del laser (*a destra*). Un laser ad anidride carbonica che abbia una potenza di 50 watt quando emette con continuità, funzionando a interruzione Q può generare ogni secondo 400 impulsi della durata di 150 nanosecondi l'uno, con una potenza che sfiora i 50 chilowatt.

chio della cavità laser con uno specchio rotante (*si veda la figura in questa pagina*).

Il laser funziona ogni volta che lo specchio rotante è parallelo allo specchio fisso sistemato sull'altra estremità del tubo laser e in questa occasione viene emesso un impulso infrarosso di 10,6 micron di lunghezza d'onda. Con una disposizione che produce l'interruzione Q, un laser ad anidride carbonica, che ha la possibilità di emettere 50 watt in funzionamento continuo, sarà in grado di emettere impulsi aventi una durata di circa 150 nanosecondi con una potenza di circa 50 chilowatt e con una frequenza di circa 400 impulsi al secondo. Gli impulsi di radiazione coerente ad alta potenza che possono essere prodotti dai laser a gas sono di particolare utilità nei sistemi di analisi fisica non distruttiva.

Una sorgente ideale di radiazione coerente è una sorgente che può essere « accordata » con continuità, cioè una sorgente la cui frequenza ha la possibilità di essere variata e controllata con continuità. Non è possibile accordare un laser molecolare ad alta potenza nel senso stretto della parola, tuttavia il numero di transizioni rotovibrazionali che possono essere scelte per l'oscillazione è notevolmente grande. Per esempio, per l'ossido di carbonio esistono circa 200 transizioni, che oscillano con lunghezze d'onda comprese fra i cinque e i sei micron; nell'anidride carbonica si possono scegliere circa 100 transizioni che sono in grado di fornire emissioni ad alta potenza con lunghezze d'onda comprese fra i nove e gli undici micron. In breve, i laser molecolari ad alta potenza permettono una notevole scelta di lunghezze d'onda discrete.

I laser ad anidride carbonica ad alta potenza, a emissione continua e a interruzione Q, possono avere molte applicazioni pratiche. Focalizzando l'emissione infrarossa coerente su una superficie di circa un millesimo di centimetro quadrato, si ottiene un'intensità di un milione di watt al centimetro quadrato se si impiega un laser a onda continua, e un'intensità di un miliardo di watt al centimetro quadrato se si impiega un laser a interruzione Q. Fasci continui, con potenze superiori al chilowatt, possono venire impiegati in semplici applicazioni tecniche come il taglio e la saldatura dei metalli. Irradiando con un laser ad anidride carbonica da un chilowatt una roccia di granito si può osservare che essa subisce uno strano « rammollimento ». L'emissione di un laser ad anidride carbonica ha una lunghezza d'onda di 10,6 micron e, sebbene sia invisibile all'occhio umano, possiede una potenza devastatrice. Le applicazioni tecniche dei laser ad anidride carbonica stanno suscitando notevole interesse nel campo industriale.

Sono tuttavia piú importanti le applicazioni che non sfruttano le proprietà termiche del laser ad anidride carbonica. Queste applicazioni includono il campo delle comunicazioni ottiche sia terrestri che spaziali. La caratteristica piú importante è rappresentata dalla bassa perdita ottica che si manifesta nelle trasmissioni attraverso l'atmosfera terrestre in corrispondenza delle radiazioni la cui lunghezza d'onda è compresa fra gli otto e i quattordici micron. L'alta efficienza e l'alta potenza dei laser ad anidride carbonica che emettono a 10,6 micron li fanno considerare candidati ideali per tali applicazioni. Il laser ad anidride carbonica è inoltre molto adatto per essere impiegato nel campo dei radar ottici, sempre per la bassa perdita nella trasmissione attraverso l'atmosfera. Un'ulteriore possibilità è costituita dal suo impiego nello studio delle interazioni ottiche di 10,6 micron di lunghezza d'onda con la materia, dato che molti semiconduttori, opachi alle radiazioni dello spettro visibile, sono trasparenti a queste lunghezze d'onda. Un'altra applicazione del laser ad anidride carbonica ad alta potenza è quella di utilizzare la sua emissione a 10,6 micron come « sostegno » per lo studio delle proprietà non lineari di nuovi materiali, con l'intento di ottenere una sorgente di radiazioni infrarosse realmente accordabile. A tal fine io e i miei colleghi stiamo eseguendo un certo numero di esperimenti assai interessanti, che includono la produzione di una seconda armonica, l'amplificazione parametrica della radiazione dell'infrarosso lontano, la produzione nei semiconduttori di coppie a due fotoni elettrone-buca, la non linearità dovuta agli elettroni di conduzione nei semiconduttori e la diffusione Raman, dovuta agli elettroni del livello Landau nei semiconduttori. Alcuni di questi meccanismi sono abbastanza importanti da permetterci di costruire un oscillatore laser regolabile nell'infrarosso. Questo laser regolabile, in connessione con il laser ad anidride carbonica a frequenza fissa, potrà essere impiegato come oscillatore locale in un sistema di comunicazioni ottiche o in un sistema radar. Inoltre, una tale sorgente di radiazioni infrarosse potrà rivoluzionare completamente la spettroscopia infrarossa. La descrizione di questi esperimenti sarà forse l'argomento di un mio prossimo articolo.

L'interruzione *Q* è un accorgimento tecnico impiegato per ottenere da un laser funzionante con continuità un'emissione impulsata. Questo accorgimento è ottenuto sostituendo uno degli specchi della cavità laser con uno specchio rotante. Il laser funziona ogni volta che lo specchio rotante (*a sinistra*) risulta parallelo allo specchio fisso che si trova sull'altra estremità del laser (*a destra*). Un laser ad anidride carbonica che abbia una potenza di 50 watt quando emette con continuità, funzionando a interruzione *Q* può generare ogni secondo 400 impulsi della durata di 150 nanosecondi l'uno, con una potenza che sfiora i 50 chilowatt.

chio della cavità laser con uno specchio rotante (*si veda la figura in questa pagina*).

Il laser funziona ogni volta che lo specchio rotante è parallelo allo specchio fisso sistemato sull'altra estremità del tubo laser e in questa occasione viene emesso un impulso infrarosso di 10,6 micron di lunghezza d'onda. Con una disposizione che produce l'interruzione *Q*, un laser ad anidride carbonica, che ha la possibilità di emettere 50 watt in funzionamento continuo, sarà in grado di emettere impulsi aventi una durata di circa 150 nanosecondi con una potenza di circa 50 chilowatt e con una frequenza di circa 400 impulsi al secondo. Gli impulsi di radiazione coerente ad alta potenza che possono essere prodotti dai laser a gas sono di particolare utilità nei sistemi di analisi fisica non distruttiva.

Una sorgente ideale di radiazione coerente è una sorgente che può essere « accordata » con continuità, cioè una sorgente la cui frequenza ha la possibilità di essere variata e controllata con continuità. Non è possibile accordare un laser molecolare ad alta potenza nel senso stretto della parola, tuttavia il numero di transizioni rotovibrazionali che possono essere scelte per l'oscillazione è notevolmente grande. Per esempio, per l'ossido di carbonio esistono circa 200 transizioni, che oscillano con lunghezze d'onda comprese fra i cinque e i sei micron; nell'anidride carbonica si possono scegliere circa 100 transizioni che sono in grado di fornire emissioni ad alta potenza con lunghezze d'onda comprese fra i nove e gli undici micron. In breve, i laser molecolari ad alta potenza permettono una notevole scelta di lunghezze d'onda discrete.

I laser ad anidride carbonica ad alta potenza, a emissione continua e a interruzione *Q*, possono avere molte applicazioni pratiche. Focalizzando l'emissione infrarossa coerente su una superficie di circa un millesimo di centimetro quadrato, si ottiene un'intensità di un milione di watt al centimetro quadrato se si impiega un laser a onda continua, e un'intensità di un miliardo di watt al centimetro quadrato se si impiega un laser a interruzione *Q*. Fasci continui, con potenze superiori al chilowatt, possono venire impiegati in semplici applicazioni tecniche come il taglio e la saldatura dei metalli. Irradiando con un laser ad anidride carbonica da un chilowatt una roccia di granito si può osservare che essa subisce uno strano « rammollimento ». L'emissione di un laser ad anidride carbonica ha una lunghezza d'onda di 10,6 micron e, sebbene sia invisibile all'occhio umano, possiede una potenza devastatrice. Le applicazioni tecniche dei laser ad anidride carbonica stanno suscitando notevole interesse nel campo industriale.

Sono tuttavia piú importanti le applicazioni che non sfruttano le proprietà termiche del laser ad anidride carbonica. Queste applicazioni includono il campo delle comunicazioni ottiche sia terrestri che spaziali. La caratteristica piú importante è rappresentata dalla bassa perdita ottica che si manifesta nelle trasmissioni attraverso l'atmosfera terrestre in corrispondenza delle radiazioni la cui lunghezza d'onda è compresa fra gli otto e i quattordici micron. L'alta efficienza e l'alta potenza dei laser ad anidride carbonica che emettono a 10,6 micron li fanno considerare candidati ideali per tali applicazioni. Il laser ad anidride carbonica è inoltre molto adatto per essere impiegato nel campo dei radar ottici, sempre per la bassa perdita nella trasmissione attraverso l'atmosfera. Un'ulteriore possibilità è costituita dal suo impiego nello studio delle interazioni ottiche di 10,6 micron di lunghezza d'onda con la materia, dato che molti semiconduttori, opachi alle radiazioni dello spettro visibile, sono trasparenti a queste lunghezze d'onda. Un'altra applicazione del laser ad anidride carbonica ad alta potenza è quella di utilizzare la sua emissione a 10,6 micron come « sostegno » per lo studio delle proprietà non lineari di nuovi materiali, con l'intento di ottenere una sorgente di radiazioni infrarosse realmente accordabile. A tal fine io e i miei colleghi stiamo eseguendo un certo numero di esperimenti assai interessanti, che includono la produzione di una seconda armonica, l'amplificazione parametrica della radiazione dell'infrarosso lontano, la produzione nei semiconduttori di coppie a due fotoni elettrone-buca, la non linearità dovuta agli elettroni di conduzione nei semiconduttori e la diffusione Raman, dovuta agli elettroni del livello Landau nei semiconduttori. Alcuni di questi meccanismi sono abbastanza importanti da permetterci di costruire un oscillatore laser regolabile nell'infrarosso. Questo laser regolabile, in connessione con il laser ad anidride carbonica a frequenza fissa, potrà essere impiegato come oscillatore locale in un sistema di comunicazioni ottiche o in un sistema radar. Inoltre, una tale sorgente di radiazioni infrarosse potrà rivoluzionare completamente la spettroscopia infrarossa. La descrizione di questi esperimenti sarà forse l'argomento di un mio prossimo articolo.

GIOCHI MATEMATICI

di Martin Gardner

Come si potrebbe comunicare con esseri extraterrestri.

Dalle profondità dello spazio, cervelli che stanno ai nostri come questi stanno a quelli degli animali mortali, intelligenze grandi e fredde e insensibili guardavano la Terra con occhi invidiosi, e lentamente e inesorabilmente maturavano progetti contro di noi.
(H.G. Wells, *La guerra dei mondi.*)

Quando, nel 1898,
Wells fu pubblicato per la prima volta, numerosi astronomi illustri erano convinti che Marte fosse abitato da creature dall'intelligenza « grande e fredda e insensibile », enormemente superiore alla nostra. L'astronomo italiano Giovanni Schiapparelli sosteneva di aver visto, nel 1877,
sottili che intersecavano la superficie del pianeta rosso. Un ricco cittadino di Boston, Percival Lowell, fu talmente affascinato dalle scoperte di Schiapparelli da abbandonare i suoi studi orientali per dedicarsi all'astronomia. Nel 1894,
vicino alla Terra, Lowell attrezzò un osservatorio personale a Flagstaff, in Arizona.

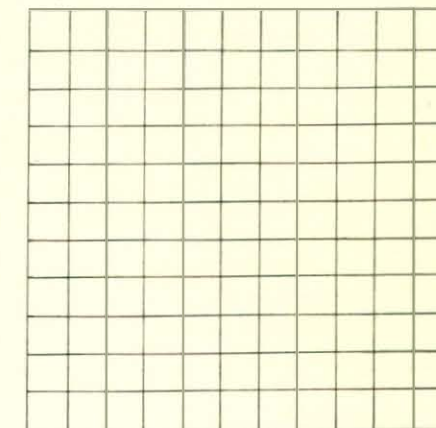
Lowell non solo vide le linee che Schiaparelli aveva chiamato « canali », ma ne registrò un numero fantastico: piú di 500. In tre libri – *Mars* (1895), *Mars and Its Canals* (1906),
the Abode of Life (1908) – Lowell sostenne che le linee da lui viste erano ampie fasce di vegetazione che costeggiavano enormi canali costruiti per irrigare i grandi deserti di Marte con le acque ottenute dallo scioglimento delle calotte polari. « Che Marte sia abitato da esseri viventi di un tipo o dell'altro » scriveva « può considerarsi sicuro; ciò che non sappiamo, invece, è di che tipo di esseri si tratti. » I libri di Lowell su Marte ebbero una grande influenza sui primi scrittori di fantascienza; i canali saltavano fuori da tutte le parti, dal breve racconto di Wells *L'uovo di cristallo* (1897) alle piú recenti storie marziane di Edgar Rice Burroughs.

La competenza di Lowell come astronomo non può essere messa in dubbio. I suoi calcoli del 1915 permisero a Clyde W. Tombaugh di scoprire, nel 1930 al Lowell Observatory, il pianeta Plutone (in latino *Pluto*), cosí chiamato perché le prime due lettere di questo nome sono le iniziali di Lowell, e le ultime due sono le prime due lettere di Tombaugh (il simbolo di Plutone, ♇, è una combinazione della P con la L). Il temperamento di Lowell, però, era piú simile a quello della sorella Amy, poetessa e fumatrice di sigari, che a quello del fratello Abbott Lawrence, prudente conservatore, presidente della Harvard University. Sebbene alcuni astronomi confermassero entusiasticamente l'osservazione di Lowell dei canali di Marte, altri, con telescopi piú potenti e con occhi migliori, non riuscivano a vederne alcuno. Anche oggi, del resto, Marte appare al miglior telescopio come un puntolino luminoso minuscolo e tremolante, e in quei rari e fuggevoli momenti in cui l'immagine è nitida è possibile che la fantasia giochi brutti scherzi. Né la fotografia può essere d'aiuto, perché la turbolenta atmosfera terrestre offusca le immagini.

L'opinione degli astronomi moderni è che Schiaparelli, Lowell e i loro seguaci siano stati vittime di illusioni ottiche dovute alla presenza, sul pianeta rosso, di chiazze irregolari, elaborate dall'astigmatismo e da una forma di autosuggestione.

Fra i pochi scienziati viventi che continuano a sostenere le tesi di Lowell il piú loquace è un chimico californiano di nome Wells A. Webb. Nel libro *Mars, the New Frontier: Lowell's Hypothesis* (1956) e in numerosi articoli e conferenze, egli ha compiuto un'interessante analisi topologica della mappa dei canali fatta da Lowell e dal suo principale sostenitore, Robert J. Trampler. Considerando le mappe di questi due astronomi come un reticolo geometrico, Webb ha determinato la percentuale di vertici a cui fanno capo tre, quattro, cinque, sei, sette e otto linee (detti rispettivamente vertici di ordine tre, quattro, cinque, ecc.). Nelle mappe di entrambi gli astronomi predominano i vertici di ordine quattro: essi rappresentano circa il 43 % dei vertici nelle mappe di Trumpler, il 55 % in quelle di Lowell. Analizzando i reticoli esistenti in natura – crepe nel fango e nelle porcellane, crepacci in antichi letti di lava, fiumi e cosí via – Webb trovò un netto predominio dei vertici di ordine tre. Analizzando invece i reticoli creati da organismi viventi, come ragnatele e piste di animali, egli trovò una netta maggioranza dei vertici di ordine quattro. Inoltre, scoprí che i reticoli topologicamente piú prossimi a quelli di Lowell e Trumpler si ritrovano in certi tipi di opere umane (linee ferroviarie e rotte di aerei e di navi). La topologia, sostiene Webb, conferma quindi la convinzione intuitiva di Lowell che i canali devono essere opera di intelligenze altamente evolute.

Le argomentazioni di Webb, però, ammettono una corrispondenza tra la superficie di Marte e le mappe di Lowell e Trumpler. Se invece queste mappe sono solo una rappresentazione di ciò che i due astronomi credevano di vedere, la loro somiglianza topologica alle linee ferroviarie è facilmente comprensibile. Le immagini televisive della superficie di Marte trasmesse dai Mariner non hanno permesso di decidere se vi siano o no delle linee sul pianeta.

Dal 1880 al 1925,
fu un grande interesse per i canali di Marte, furono avanzate numerose proposte per stabilire contatti con i marziani. Per esempio, si suggerí di costruire un potentissimo riflettore con

La matrice 11 × 11.

1. A.B.C.D.E.F.G.H.I.J.K.L.M.N.P.Q.R.S.T.U.V.W.Y.Z.

2. AA,B;AAA,C;AAAA,D;AAAAA,E;AAAAAA,F;AAAAAAA,G;
AAAAAAAA,H;AAAAAAAAA,I;AAAAAAAAAA,J.

3. A K A L B;A K A K A L C;A K A K A K A L D.A K A L B;B K A L C;C K A L D;
D K A L E.B K E L G;G L E K B.F K D L J;J L F K D.

4. C M A L B;D M A L C;I M G L B.

5. C K N L C;H K N L H.D M D L N;E M E L N.

6. J L AN;J K A L AA;J K B L AB:AA K A L AB.J K J L BN;J K J K J L CN.
FN K G L FG.

7. B P C L F;E P B L J;F P J L FN.

8. F Q B L C;J Q B L E;FN Q F L J.

9. C R B L I;B R E L CB.

10. J P J L J R B L S L ANN;J P J P J L J R C L T L ANNN.J P S L T;J P T L J R D.

11. A Q J L U;U Q J L A Q S L V.

12. U L NWA;U P B L NWB;AWD M A L NWD L D P U.V L NWNA;
V P C L NWNC.V Q J L NWNNA;V Q S L NWNNNA.J P EWFGH L EFWGH;
S P EWFGH L EFGWH.

13. GIWIH Y HN;T K C Y T.Z Y CWADAF.

14. D P Z P NWNNIB R C Q C.

Il messaggio interplanetario di Ivan Bell.

cui dirigere verso Marte gli impulsi luminosi di un messaggio in codice, oppure di sistemare una serie di luci su una grande area, in modo da disegnare uno schema del teorema di Pitagora visibile dai telescopi marziani. Si discusse a lungo circa i collegamenti radio: se inviare una serie di « bip » che rappresentassero i numeri interi (bip; bip, bip; bip, bip, bip; ecc.) o alcuni semplici calcoli aritmetici come due più due uguale a quattro. Nel 1900, Nikola Tesla dichiarò di aver ricevuto segnali radio da esseri intelligenti abitanti su Marte, e vent'anni più tardi anche Guglielmo Marconi annunciò qualcosa di analogo. Gli spiritisti erano in frequente contatto con intelligenze del pianeta rosso. Il caso più interessante fu quello di una medium svizzera, Hélène Smith, la cui strana storia è raccontata nel libro *Dalle Indie al pianeta Marte: studio di un caso di sonnambulismo con glossolalia* (1905) dallo psicologo svizzero Théodore Flournoy. Quand'era in trance, Hélène sembrava essere sotto il controllo marziano, e parlava e scriveva un complesso linguaggio.

Oggi che siamo in procinto di far atterrare su Marte sonde per l'esplorazione del pianeta, e ci aspettiamo di trovare, al più, vegetazione a basso grado di sviluppo, l'interesse per le comunicazioni extraterrestri si è rivolto verso i pianeti di altri sistemi solari. Nel 1960 il Progetto Ozma non riuscì a rivelare alcun messaggio radio proveniente dagli spazi esterni dopo parecchi mesi di ascolto in prossimità della frequenza a cui emettono gli atomi di idrogeno libero (per varie ragioni pare che questa frequenza, con una lunghezza d'onda di 21 cm, sia l'ideale per le comunicazioni interstellari). Malgrado ciò, l'interesse sia per la ricerca sia per la trasmissione di tali messaggi continua a essere vivo e si sono compiuti, e si stanno compiendo, numerosi lavori miranti a trovare il metodo migliore per scambiare informazioni con altre culture, una volta che si sia riusciti a stabilire il contatto. Questo affascinante problema è l'esatto opposto di quello di escogitare codici per messaggi segreti. In questo caso, infatti, la trasmissione di informazioni deve avvenire in modo tale da rendere il più difficile possibile, per chi non conosca la chiave, la comprensione del messaggio. In un codice interstellare, invece, l'obiettivo è comunicare con intelligenze che nulla sanno del nostro linguaggio, per cui si deve rendere il più facile possibile il messaggio stesso.

La maggior parte delle pagine di *Interstellar Communication*, una raccolta di scritti di autori diversi a cura di A.G.W. Cameron (1963), tratta questo problema. Tutti gli esperti sono d'accordo che i messaggi devono cominciare dall'aritmetica più elementare. Alcuni postulano che tutti i tipi di esseri intelligenti conoscano le unità, e che le leggi aritmetiche siano uniformi nella galassia. Non è però possibile ammettere che tutti i metodi per simbolizzare i numeri – per esempio la nostra notazione posizionale a base 10 – siano universali. Sarebbe follia, per esempio, pensare di attirare l'attenzione di esseri extraterrestri trasmettendo lo sviluppo decimale di π, in quanto se questi esseri, come è molto probabile, usano un altro sistema di base, il nostro π gli apparirebbe come una serie di simboli casuali. Hans Freudenthal, un matematico olandese, ha inventato un complicato linguaggio chiamato « Lincos » (da « Lingua Cosmica ») che parte dall'aritmetica e dalla logica elementare, si sviluppa fino alla matematica più complessa e giunge infine alla comunicazione di ogni conoscenza umana. Il primo volume della sua opera è stato pubblicato in Olanda nel 1960.

La maggior parte degli sforzi di Freudenthal si rivelerà forse di scarso interesse in quanto è molto facile trasmettere immagini usando un semplice codice di due simboli. Questo sistema non richiede che gli esseri che ricevono il messaggio abbiano occhi fotosensibili, ma solo che siano in grado di registrare la forma delle cose; le nostre immagini possono allora venire da essi tradotte in qualunque forma sensorialmente percepibile, che costituisca il loro modo abituale di « vedere » le cose. Forse il sistema più semplice per trasmettere una forma è un messaggio a due simboli, contenente le istruzioni per la scansione di tutte le celle di una matrice, rettangolare o quadrata: un simbolo vuol dire che la cella è piana, l'altro che essa è vuota. Questa tecnica è normalmente usata per trasmettere le immagini via radio, e costituisce la base della scansione televisiva. Consideriamo, a titolo di esempio, il seguente messaggio di 121 simboli:

1 0 0 0 1 0 0 0 0 0 0
1 1 1 1 1 0 0 0 0 0 0
0 1 1 1 1 1 1 1 1 1 1
0 1 1 1 0 1 1 1 1 1 1
0 0 0 0 0 1 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 0
1 1 1 0 1 0 1 0 1 1 1
1 0 1 0 1 0 1 0 1 0 0
1 1 0 0 1 0 1 0 1 1 0
1 0 1 0 1 0 1 0 1 0 0
1 1 1 0 1 1 1 0 1 1 1

Questi 121 simboli suggeriscono la matrice 11 × 11 mostrata nella figura a pagina 93. Se il lettore esamina ogni casella da sinistra a destra e dall'alto in basso, annerendo le celle indicate con 1, troverà l'immagine e il nome di un ben noto animale. È chiaro che, una volta afferrato il principio della scansione delle immagini, si può progredire nelle comunicazioni con straordinaria facilità e rapidità.

Poiché sono necessari centinaia o migliaia di anni affinché un messaggio percorra la distanza fra la Terra e un pianeta di un altro sistema solare, è evidentemente impossibile fare lunghe chiacchierate come se si fosse al telefono. I messaggi dovrebbero iniziare con qualcosa che possa attirare l'attenzione – i numeri interi o una serie di numeri primi – seguito da semplici regole aritmetiche che portino velocemente alla scansione delle immagini, e quindi alla trasmissione di informazioni di vario tipo. Ma quali informazioni vanno trasmesse per prime? Il problema non è semplice. Si potrebbe pensare che le prime informazioni da trasmettere siano riferite a quelle cose che i fisici chiamano « osservabili » – informazioni cioè derivate dai nostri sensi, spesso aiutati da strumenti d'osservazione relativamente semplici, come telescopi e microscopi. Ma supponiamo che gli abitanti del pianeta X abbiano sviluppato soprattutto una facoltà di percezione che noi non abbiamo mai posseduto; per esempio, che percepiscano sensorialmente le forze di natura magnetica o certi tipi di radiazioni a noi sconosciuti. Sul pianeta X, le nostre forme del mondo, derivate dalle nostre osservabili, potrebbero avere minor significato di informazioni relative a certe « non osservabili » come elettroni, protoni e neutroni. Se ciò fosse vero, gli abitanti del pianeta X non avrebbero difficoltà a capire una descrizione della Tavola periodica degli elementi, mentre forse sarebbero imbarazzati di fronte alla rappresentazione di una casa o di un albero. Da un certo punto di vista, i colori, le forme, i suoni del nostro mondo sono fatti fondamentali, e gli elettroni una vaga astrazione. Al contrario, la struttura matematica dell'atomo di elio può risultare più universalmente comprensibile del colore, del sapore e della forma di una mela, non perché è improbabile che esistano mele su un altro pianeta, ma piuttosto perché altri esseri possono percepire il loro universo con sensi che poco hanno in comune con vista, udito, olfatto, gusto e tatto. Entità teoriche come le particelle e i campi elettromagnetici possono risultare più facilmente comprensibili agli esseri extraterrestri degli odori e dei suoni del nostro mondo.

Nel 1960 l'inglese Ivan Bell, che insegnava la madrelingua a Tokyo, lesse degli articoli sul Progetto Ozma. Per divertire gli amici immaginò un semplice messaggio interplanetario di 24 simboli, pubblicato su « The Japan Times » il 22 gennaio 1960, chiedendo ai lettori di cercare di decifrarlo. Alla redazione del giornale giunsero quattro soluzioni complete.

Il messaggio di Bell è riportato nella pagina a fronte. Esso è più facile di quanto possa sembrare, e i lettori sono invitati a decifrarlo. I 24 simboli sono le lettere dell'alfabeto dall'A alla Z (escluse la O e la X). Ogni simbolo può venir trasmesso, presumibilmente, mediante una particolare combinazione di bip, ma per i nostri scopi ciò non ha interesse. La punteggiatura non fa parte del messaggio, ma indica intervalli di tempo; le lettere adiacenti sono trasmesse con brevi pause tra l'una e l'altra. Uno spazio tra le lettere significa una pausa più lunga, e le virgole, i punti e virgola e i punti rappresentano a loro volta pause progressivamente più lunghe. Fra paragrafi diversi, numerati per comodità del lettore, sono lasciati intervalli di tempo ancora più lunghi. I numeri non fanno parte del messaggio. Il messaggio dovrebbe essere inequivocabilmente chiaro per esseri di ogni sistema solare, a eccezione dell'ultimo paragrafo, che è un poco ambiguo (anche se decifrato correttamente esso potrebbe venir compreso solo dagli abitanti di uno dei pianeti del nostro sistema solare).

I due alberi coincidenti, nella « foresta » di 12 alberi con 7 vertici, sono quelli indicati con 5 e 7.

Da qualsiasi posizione di partenza per il solitario dell'orologio è perdente. Il suo diagramma non è un albero; non solo non è connesso, ma una parte di esso contiene un circuito.

L'illustrazione in questa pagina mostra come disegnare alberi di lunghezza minima e congiungenti i vertici di un quadrato e di un pentagono regolari. Gli angoli indicati dal punto pieno sono di 120°. Si noti che le due diagonali di un quadrato forniscano un albero di economia (di lunghezza $2\sqrt{2} = 2{,}828$... nel caso di un quadrato di lato unitario), ma l'albero indicato misura $1 + \sqrt{3} = 2{,}732$... (sempre se il lato del quadrato è unitario). Una dimostrazione senza calcoli del fatto che si tratta proprio dell'albero di lunghezza minima è data in *100 Problems in Elementary Mathematics*, di Hugo Steinhaus (Basic Books, 1964), problema 73. L'albero minimo di un pentagono di lato unitario è 3,

Il minimo albero interno a un triangolo equilatero ha un quarto punto nel centro del triangolo. Gli alberi minimi per poligoni di sei o più lati sono costituiti semplicemente dal perimetro con un lato in meno. Per quanto riguarda il problema generale di trovare gli alberi minimi che collegano *n* punti del piano – e una tecnica per trovare tali alberi sfruttando la tensione superficiale di una pellicola di acqua e sapone – si veda *Che cos'è la matematica?*, di Richard Courant e Herbert E. Robbins, Boringhieri, 1964,

Gli alberi di lunghezza minima.